DOCUMENT RESUME

ED 074 519 CS 200 453

AUTHOR	Battersby, James L.
TITLE	Typical Folly: Evaluating Student Performance in Higher Education.
INSTITUTION	National Council of Teachers of English, Urbana, Ill.
PUB DATE	73
NOTE	72p.
AVAILABLE FROM	National Council of Teachers of English, 1111 Kenyon Road, Urbana, Ill. 61801 (Stock No. 05354, $2.25 non-member, $2.00 member)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Academic Performance; Academic Records; *Achievement Rating; Evaluation Criteria; *Grades (Scholastic); Grading; Pass Fail Grading; Standardized Tests; *Student Evaluation

ABSTRACT

This study investigates various methods of evaluating student performance in higher education. Chapter one discusses some of the possible reasons that grades given to students seem to be consistently higher than in the past. The author argues that this tendency may be due to increasing dissatisfaction with the conventional grading system. Chapter two discusses the shortcomings of letters of recommendation and standardized national exams as predictors of future performance. Chapter three argues that grades are unsuited to signify the quality of student work, inadequate as predictors of success, and inappropriate as means of discriminating among different students' achievements. Chapter four discusses some of the many diverse and unstable criteria used in assigning grades. Chapter five examines some alternatives to the grading system but finds that most of these suffer from the same weaknesses that affect grades. Chapter six proposes written evaluations of student achievement as a possible substitute for grades and discusses the possibility of combining a credit/no record system with a written evaluation system. A bibliography is included. (Author/DI)

ED 074519

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

Typical Folly:

Evaluating Student Performance in Higher Education

James L. Battersby
The Ohio State University

CS 200453

NATIONAL COUNCIL OF TEACHERS OF ENGLISH

NCTE COMMITTEE ON PUBLICATIONS Robert F. Hogan, Blyden Jackson, Robert E. Palazzi, Bette J. Peltola, Eugene C. Ross/STAFF EDITOR Linda Jeanne Reed/STAFF DESIGNER Norma Phillips Meyers

CONSULTANT READERS Walker Gibson, University of Massachusetts, Amherst; Richard Ohmann, Wesleyan University, Middletown, Connecticut

Library of Congress Catalog Card Number 72-94889
ISBN 0-8141-0535-4
NCTE Stock Number 05354

Copyright ©1973 by the National Council of Teachers of English.
All rights reserved. Printed in the United States of America.

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY
National Council
of Teachers of English
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

Table of Contents

Introduction	vii
1. San Francisco State as Everycollege	1
2. Loan Agents and Soothsayers	6
3. Probability Judgments and Class Tendencies	15
4. Confusion Compounded	24
5. Alternatives to Grades	29
6. A Proposal: Written Reports	35
Bibliography	55

Introduction

As the title indicates, the principal subject of this book is the evaluation of student performance in higher education. I begin with the assumption that grades, the traditional means of signalizing academic achievement, are rising, that instructors throughout the nation are awarding higher grades today than they have in the past. The assumption is, I think, well-founded; all the statistical evidence I have seen suggests that grades are on the rise nationally. The causes of this phenomenon are undoubtedly many and various, but chief among them, judging on the basis of my own experience and that of my colleagues, is an increasing dissatisfaction with the grading system as a means of designating the nature of student achievement. After weighing what the various grade signs can possibly *mean in themselves* against what certain of those signs may *come to mean* to the student and to those who will interpret them for such purposes as determining whether the student will stay in school or whether he will be accepted into graduate school, many instructors assign "tolerable" grades to less than mediocre work. Recognizing that final grades convey very little educational information to students but speak directly and meaningfully to those who, after looking at grades in the aggregate, have to do something about what they see—recognizing, in short, that the logical nexus is from grades to administrators, selection committees, prospective employers, and

so forth—instructors with increasing regularity assign grades which directly reflect an awareness of the extra-course, extra-educational significance of their "academic" grade decisions.

Whatever the causes, this general rising of grades creates interesting evaluation problems. When grades cease to discriminate effectively among students (an inevitable consequence of rising grades), administrators, selection committees, and admissions officers are obliged to look to other evidences of academic and intellectual ability for their criteria of differentiation. As the determinate significance of grades diminishes, the importance of letters of recommendation and standardized national exams increases. Thus, important decisions relating to the future activity and mobility of students are more and more frequently based on evidence further and further removed from the actual, demonstrated academic achievement of students. In chapter 2, I discuss at length the radical shortcomings of letters of recommendation and standardized national exams as predictors of future performance possibilities, as conveyors of information genuinely useful to those responsible for probability decisions.

In order to understand the particular focus of this book the reader should know that the greater part of my argument deals with the problem of how to make probability judgments more reliable than they presently are without relying on evidence only tenuously related to what a student accomplishes in a long succession of endeavors. I have assumed that grades, like most other systems of evaluation, are more administrative conveniences than educational necessities, that the information upon which intellectual growth and development depend is not conveyed to the student by whatever sign is attached to his work, and that, in general, the student determines how well he is doing by attending to the critical and analytic comments accompanying the grade sign, by attending to how he is treated (post exam, quiz, paper, experiment) by his instructor, by measuring himself against what he conceives to be possible, against instructorial expectations, against estimable standards of excellence, and so forth. (Even the "self-justifying" grade placed on the so-called objective test is, from the standpoint of communication, superfluous, since the educationally meaningful message is transmitted by the marks indicating "wrong" answers.) Briefly, I have assumed that the signs of student achievement submitted to the registrar's office at the end of each term are desiderata of administrative and judicial agencies, not of students and teachers. Clearly, there is no imperative implicated in the teaching-learning process calling for fulfillment in a system of evaluation based on a prescribed set of three, four, five, six, seven, or eight discrete signs. Nevertheless the fact remains that decisions of great significance and important

personal consequence will continue to be made about students, regardless of whether instructors submit reports of student progress and achievement to the registrar's office, and it is an abiding concern with this administrative fact which in large measure determines the nature, direction, and end of my argument.

Having argued against judgments informed by the evidence presented in commendatory letters and national exams, I suggest in chapter 3 that the most reliable information on which decision makers can draw in their efforts to anticipate the future is the pattern of achievement that a student makes manifest in the process of meeting specific academic problems over an extended period. The guiding assumption here is that only when the nature and quality of past performance is known in some detail can we begin to determine how the qualities of mind implicated in past effort may be expected to insinuate themselves into the future. At the present educational moment the evaluation system which is best suited to provide immediate access to a student's demonstrated on-the-job academic performance is the grading system, the discriminative power of which a multiplicity of factors (examined in the first two chapters) are conspiring to undermine. Chapter 3 sets out to demonstrate that, even in a non-inflationary grade market, grades are radically unsuited to signify the nature or quality of student work and that, even if the opinions about their values were uniform, we would still have no way of knowing what we most need to know about the achievement actually realized by the student. At best, all probability judgments are guesses, attempts to subsume what is unknown under some known class; grades, by their very nature, cannot disclose information sufficiently precise to inform differential judgments with any significant degree of probability. In short, each grade allows too many and excludes too few possible meanings; the reader of the sign simply has no way of knowing whether his interpretation corresponds to the instructor's intended meaning. Thus, even under optimum conditions (i.e., when the values assigned to grades are commonly shared) the explanatory and discriminatory power of grades is too severely limited to serve the purposes of those who wish to discriminate among genuinely differentiable achievements.

Actually, of course, there is no consensus regarding the values of individual grades, and in chapter 4, I briefly examine some of the marvelously diverse criteria to which student work must adjust and conform if it aspires to instructorial approval, noting that it is this instability of criteria which contributes substantially to the production of safe, timid, non-grade-point-average-threatening work.

Perhaps no educational topic has generated more discussion than the

grading system. To many, grades are the most odious little vermin ever suffered to take up residence in the house of intellect. Others, unwilling to carry criticism to the heights of fulmination, are content to abuse grades and, after bemoaning their inability to devise a better system of evaluation, to accept them grudgingly as necessary evils, as the ineluctable concomitants of mass education. Even those who support grades tend to emphasize their convenience, not their educational value. What some critics despair of finding, however, others have discovered within easy reach, namely, "adequate" replacements for the outmoded grade mechanism. Aware of the growing dissatisfaction with grades and, presumably, of the inherent limitations of the grading system, many schools are experimenting with a variety of alternative evaluation systems. Chapter 5 argues that the "new" systems (including the phrase system–"high distinction," "distinction," etc.–and pass/fail) are little more than ill-disguised variations of the old grading system and that virtually all of the "innovative" programs suffer from the same weaknesses that afflict grades. Evaluation and prediction continue to be grounded in mystery and uncertainty, because they distinguish complex and various achievements by means of a prescribed, limited, and unchanging set of discrete signs or terms.

In a real sense the rationale for chapter 6, in which I discuss the merits of an evaluation system based on written reports, is to be found in the antecedent chapters; to my mind, the problems raised throughout the book come to proper resolution in detailed evaluations which would form the substance of the students' academic records and be the principal resource of reviewers responsible for making probability judgments concerning students. It is perfectly clear that between the analysis and criticism of student work (aspects of education absolutely essential to the student who hopes to free himself from the cramp and confinement of intellectual isolation and to define his efforts in relation to those generally acknowledged standards of excellence intrinsic to the discipline in which he has elected to work) and the submission of signs of achievement to the registrar's office for recording in the student's academic bank account, there is a large, conspicuous space which cannot be bridged by the evaluation systems considered earlier. In essence, the written evaluation proposal is designed to bridge this gap by making the process of reviewing student "credentials" depend upon the accumulated evidence of student performance, evidence which criticism and analysis supply. The specific merits of this system are discussed at length in the text. In the concluding section of the chapter, I attempt to respond to as many practical objections to the proposal as I can anticipate. Also, since any dramatic change in evaluation practices will inevitably have an impact on aspects of

higher education ostensibly and genuinely unrelated to evalution, I have given some attention, in passing, to a few of those educational reforms clearly implicated in (and perfectly consonant with) the evaluation proposal described in chapter 6. In chapter 6, I also discuss the credit/no record system, a system vigorously supported by many students and educational critics today, and I explain how credit/no record could most effectively be used in conjunction with an evaluation system based on detailed reports.

The proposal with which this book concludes is, I think, both intellectually sound and eminently practical, if not highly original. At any rate, the particular problems to which the argument of this book has given prominence find, it seems to me, their adequate and natural resolution in a system of evaluative reports. And although considerable emphasis has been placed on how the proposal can bring the process of evaluation into conformity with the results of criticism and analysis and the process of reviewing into contact with the nature of on-the-job student performance, the proposal, finally, is valuable only because it does not, in its operation, tend to get in the way of our efforts as students and teachers to concentrate attention on the beauty, power, grace, truth, and human significance of the material which we have an opportunity to examine, study, enjoy, or create.

San Francisco State as Everycollege

One year after San Francisco State College was the scene of general chaos and an extended strike which virtually closed the campus down for a period of several weeks, the school was forced to confront an issue that could be raised at every college in the nation, with perhaps a handful of exceptions. A report was issued in the fall of 1969 indicating that students were achieving higher grades than at any time in the past, the particularly alarming fact being the high percentage of A's granted to students at the end of the semester of campus upheaval. The facts were presumably indisputable; no one chose, at any rate, to challenge their validity. What cannot be denied, however, can often be explained or justified, and administrators and teachers responded with alacrity to the implicit demand for explanation in the report, the immediate results being implications of negligence on the one hand and assertions of extraordinary powers of intellect on the other. In brief, the president of San Francisco State, S.I. Hayakawa, stated that rising grades could be attributed to a variety of causes, including increased graduate enrollment and the presence of rising numbers of public school teachers in the classes. Allegedly, Hayakawa also declared that the high scores reflected the misguided generosity of instructors who were willing to relax rather than to enforce academic standards in an effort to avoid penalizing students whose demands they tacitly, if not openly,

supported. A spokesman for the College Teachers' Union, responding to the accusation, charged that Hayakawa's remarks impugned the integrity of the teachers and debased the achievement of the students. He also added to the list of causes for rising grades the rapid increase in the number of students who, because of competent performance in junior colleges, had been admitted to San Francisco State at the upper-division level. No factor stands out more prominently in the statements of the representative of the union, however, than the putative intellectual superiority of the current crop of students to those who formerly occupied the groves of academe.

These responses are, I think, both natural and predictable, but what is especially unfortunate in all this wrangling is the fact that a serious issue is lost in a cloud of charges and countercharges. Nothing is perhaps more generally true than the assertion by the union spokesman that grades have exhibited a tendency to rise for several years now throughout the country, but to attribute rising grades to the intellectual superiority of the present generation of students is, at best, to explain a complex phenomenon in terms of an unsubstantiated cause, and *that* the least compelling and the most questionable. Any impartial appraisal of the facts would include among the contributing causes, in addition to those already enumerated, a growing reluctance among faculty members, particularly in the humanities and social sciences, to apply letters or numbers to grade sheets with some degree of confidence in the efficacy of such signs to designate adequately the nature of an individual's academic performance. Lacking any conviction that the sign conveys a universally sharable meaning or that it has been arrived at by relying on uniform principles of judgment, and knowing further that (in a period when selective service agencies, scholarship committees, and "benevolent" foundations–not to mention, for the moment, graduate schools and corporations–watch for grade-point averages with a tireless vigilance) the grade assigned to a student may provoke a personal calamity grossly incommensurate with the academic implications of the sign or may carry a meaning bearing absolutely no relation to that intended, many instructors, usually after much thought, assign high grades in an effort to prevent meaningless signs from attaining a meaning deleterious to the student in the world beyond the classroom. More and more instructors are unwilling to grant D's or F's because they are convinced that the uses to which such grades will be put are unrelated to their academic significance and inconsistent with the academic needs of the students.

The hard-line response to the conscientious instructor's dilemma is clear enough: "The uses to which your grades will be put are none of your concern; your job is to evaluate students according to the highest criteria

of excellence established by your discipline. You are not running a little league team which allows everyone, regardless of ability, a turn at bat; in academia, as in the world at large, the name of the game is survival of the fittest. You undermine the integrity of your subject and of academic freedom itself when you assume responsibilities not specifically listed in the description of your publicly supported job." Indeed, we all know how the academician should respond to external forces encouraging him to compromise his principles and yield to extra-academic exigencies; like Prometheus, he should adamantly refuse to submit to either threat or enticement. We have seen the movie, in all its variations, scores of times. (In one version, the school basketball star, who brings thousands of fans to the gym and dollars to the treasury, concentrates on the hoop, not on Philosophy 101. Consequently, he is put on probation, and left hanging in the balance are the big game and the building fund. Neither the threat of firing nor the lure of a full professorship will induce the philosophy teacher to lower his standards. In this conflict of good versus good, everything, of course, is resolved to everyone's satisfaction. On the night of the big game the star, while wriggling into his jock and uniform in the locker room during the first period, briefly but cogently reviews the history of Western thought from the pre-Socratics to Wittgenstein, with special emphasis on Kant and the categorical imperative. The professor's standards are maintained; the star goes on to win the game and the school sweetheart; the construction of new buildings is imminent; and we are left with the understanding that after a stunning career in professional basketball the star will begin graduate study under the direction of his old philosophy teacher and, finally, will write a brilliant monograph on games theory, the outlines of which are briefly adumbrated on his way to the showers.)

Instructors today, of course, are not dealing with the peculiar tribulations of athletic superstars inhabiting the mythical kingdom of Celluloid but with what amounts to the ontological status of each of their students. And ontological status is not an entirely inappropriate term. The nature and quality of subsequent life is often written large in the student's grade-point average,[1] and we have come to define one another, in and out of academia, almost exclusively in quantitative terms. ("He's a 2.5 sort of guy." "He's a 35 thou a year man.") The dismal and inescapable fact is that the proper functioning of our society is dependent upon the application of production criteria, requiring quantitative measures, to

[1] As explained below, a high grade-point average is a necessary but not a sufficient condition of academic and social mobility, and this is so largely because grades continue to rise generally.

virtually every aspect of our lives. In conformity with the imperatives inherent in the values and priorities of the society at large, the instructor provides evaluations that can be quickly tabulated and attaches to young men and women, passing from anxiety to uncertainty, cards of identity which often bear the unmistakable likeness of the albatross. It is absurd to invite the instructor to make his decisions in the hermetically sealed vacuum of intellectual criteria alone, when the ether of the general society permeates every closet. Moreover, some disregard for the extramural consequences of his intramural decisions would perhaps be warranted, if now, as in the not too distant past, a diversity of acceptable and decent alternatives was available to those either unwilling or unable to make academic performance square with rigid, intellectual criteria. It is a commonplace, but nevertheless true statement that there are depressingly few opportunities available to bright, enterprising, creative men and women outside the academic and corporate structures of American society, and every year we know with increasing certainty that credentials precede existence. Every aspect of contemporary education is informed with production criteria and propelled by quantitative measures. And grades are but a particularly obvious example of the general quantification of academic life.

What happens, of course, when grades rise generally and few students sink is precisely what happens when there is too much wealth in the economy, too much money in circulation—inflation and devaluation of the coin of the realm. Academic currency comes in several denominations (chiefly A, B, C, D, and F), and when there are many more A's and B's in circulation than D's and F's, then, of course, the worth of an individual's academic bank account is diminished—his plenty makes him poor. And when the signs that were designed to provide those responsible for personnel decisions with "objective" bases for choices no longer allow even crude discriminations among candidates, new "objective" determinants of worth must be consulted; but since no one surrogate has achieved universal authority, the way is cleared for a rich variety of more or less respectable pretenders to objective status, the one emerging as preeminent in any given case often being dependent upon the personal (and perhaps capricious) predilections of specific personnel managers or selection committees.[2] Thus the whole process of civilized deceit, in which the candidate consents to trust his future to indeterminate and generally meaningless signs that reviewers consent to accept as both determinate and meaningful, is undermined, and the individual candidate for preferment is faced with the

[2] By now, most of the alternative authorities have been established, and I discuss several of them below.

difficult task of trying to be a hustler in a game for which there are no clearly defined rules that he can use in mapping out a strategy prior to a decision.[3] As the decision-making process is preoccupied with evidence further and further removed from the student's qualifications for the position for which he is a candidate, final decisions must increasingly be based on irrelevant, idiosyncratic, or inadequate criteria. High grades become less useful but, unfortunately, not less necessary, since they cannot be discounted unless they are competitively high. A system of signs designed to provide quick and convenient access to differentiable varieties of achievement comes finally to be the least important factor in the actual process of discriminating among students.

[3] Although the resourceful and ambitious student soon devises strategies for dealing with the new games, the fact is that when advancement ceases to be based principally upon grades, the extent to which the student has control over his future is severely diminished. For example, the student has very limited control over what an instructor will include or omit in a letter or recommendation, and once it is written the student is powerless to change or devalue it. And scores on national exams cannot be corrected by a determination to work harder in the future.

Loan Agents and Soothsayers

2

Here it would perhaps be useful to distinguish between grades which function as indicators of future performance probabilities and grades which function as guarantors of an extension on the educational lease, noting at the outset that grades perform only the latter function efficaciously. A low grade in a particular course may bring the student's grade-point average beyond the toleration threshold of his sponsors (i.e., parents, those in charge of the distribution of scholarship and fellowship funds, or, to raise a recurrently significant issue, the members of his selective service board) and hence precipitate the immediate foreclosure of his educational mortgage. As a result, grades tend to rise in direct proportion to the probable severity of the consequences attendant upon the assignment of a low or failing grade. The implicit educational imperative in a low or failing grade is, generally, that a student be required to take additional work or to repeat material in a particular subject matter, and many instructors are, I think, justifiably reluctant to permit a sign calling for the correction of inadequacies or the elimination of deficiencies to become the basis of a decision which will preclude the realization of the pedagogic directive by removing the student from school, thus undermining at least part of the purpose of the grade. Even if an instructor were convinced that a failing grade should legitimately be interpreted as an indication of an individual's

radical incompetence or native inability to achieve levels of performance within minimal limits of acceptability, he more and more regularly would rather pass such a student than propel him toward a society that is prepared to embrace him with slightly less enthusiasm than that with which a healthy man embraces a leper. Again, the fact is that there are fewer and fewer viable alternatives available to the person without credentials.

The gentlemanly C, attesting at one time to undistinguished but acceptable proficiency, has become the omnibus category into which performance ranging from the tolerable to the grossly inadequate may be put.

However, because the overseers of fellowship and scholarship funds and the distributors of selective service classifications are busy men who must make rapid decisions on the basis of readily accessible information, grade-point averages do in fact serve the purpose of keeping many students in school, not because the grades permit fine discriminations among students, but precisely because they do not and because they remove from decision making any serious consideration of achievement and—incidentally—any doubt or guilt. Whether the student stays in or leaves school, the reviewers can, after looking at averages and referring to tables, confidently assert that the decision is out of their hands.

Perhaps it may be forcibly argued that most instructors do not indulge quixotic delusions of grandeur when they assign grades and, consequently, do not for a moment think that their marks could figure prominently in the subsequent direction of a student's life, but I suspect that the argument loses its force in its formulation and that its validity is enforced, if at all, by wish, not by experience or practice. The dismaying fact is that the principles informing the structure of American education today with purpose, direction, and form are derived from the needs, priorities, and ends of the extramural society, and even so ostensibly insular and local a matter as grades is made, with or without the consent or support of academicians, to pledge allegiance first to extra-academic goals and then to the educational needs and capacities of students. Moreover, extra-academic ends are served even when educators consciously or unconsciously resist those ends by raising grades (thereby preventing the possible consequences from taking immediate effect), since the educational response does nothing, finally, to eliminate the social or administrative facts provoking the response. Consequently, academicians increasingly tend to complicate their educational decisions with considerations generated by a severely restricted set of extra-academic realities and, hence, to assign grades which are responsive to or which at least reflect an awareness of those realities.

As a means to an extension on the lease, grades often do their job efficiently and provide immediate benefits to individual students. As

predictors of future performance and as indicators of present levels of performance, however, grades are virtually useless. When many come a courtin' disguised as noblemen, it is impossible to use sartorial standards to isolate the true princes from the pretenders. Thus, when grade-point averages fail to provide the basis for valid discriminations, personnel managers and selection committees (in and out of departments and schools) must resort to other criteria of differentiation.

One of the criteria most frequently called into service by admissions officers or selection committees is, of course, the unofficial but ubiquitously recognized rank of the school at which the candidate achieved his grade-point average. The ruling hypothesis is that a student graduating from, say, Harvard, is probably better qualified for graduate study than a candidate graduating from a less prestigious school. An additional refinement within the same system of judgment is that the candidate who attains a 3.0 average at High Priority College is a better risk than a candidate from Low Priority College with a 3.9 average. The assumptions underlying these distinctions can perhaps be defended, but it is necessary to remember that the decisions which necessarily follow must confirm and not test the validity of the assumptions. The occasional Low Priority graduate who, once admitted, transcends expectations by performing with distinction does not contribute to the overturning of the hypotheses governing graduate admissions. His presence is something of an anomaly; consequently, his performance provides no useful statistical data. The assumptions, in short, are self-validating. Allowing for some exaggeration, the point remains that at this stage the specific nature of academic achievement is less important to reviewers than the prestige of the school awarding the degree.

Another commonly tapped resource is the letter of recommendation. This, too, is a stratified resource involving layers of respectability, for it is clear that the attention paid to a particular letter is largely determined by the national reputation of the writer; the greater the reputation, the more seriously the letter is considered. (Since the "prestigious" professor is usually—or frequently—employed by a "prestigious" school, the candidate from such a school comes to the reviewing process with an impressive accumulation of incidental benefits.[4]) The "successful" undergraduate,

[4] Of course, the prestige of both school and instructor is generally determined by quantitative measures and production criteria. Prestige is too frequently, as applied to the instructor, a title signifying visible production (i.e., books, articles) and, as applied to the institution, a title signifying the accumulation of those whose production is visible. The bright man who simply works and thinks hard has little chance of establishing a lasting affiliation with the "prestigious" school.

who learns nothing so well in four years as the subtle intricacies of gamesmanship, will, of course, attempt to secure a letter from the most "influential" professor, regardless of how casually the professor knows or how dimly he recalls the student's actual work. It's just plain good business to obtain a recommendation from someone whose name alone is worth more than an accurate, perceptive, detailed evaluation from a young, unpublished professor.

Nevertheless, the letter of recommendation is a potentially valuable source of information, since the writer is obliged to comment on the nature and quality of a student's work, indicating in writing the particular congeries of habits, skills, and powers that a student brings to intellectual problems. In actual fact, the letters are, for the most part, unreliable as intellectual character witnesses. No man likes to think that his energy is expended without effect, that his testimony is without influence; hence, the letter writer, knowing that many are being strongly supported, often tends to neglect the student in an effort to distinguish his commendation. Even though the student is depicted as the local dwelling place of the most respected intellectual virtues, the silent reference is generally to the writer, whose praise is more or less covertly generated by the maxim, "It takes one to know one–I instill in others what I praise in them." (Surely, self-praise is the sincerest form of flattery.) To all of this panegyric the reviewers bring, of course, a skeptical eye. The common practice among reviewers, I suspect, is to divide all praise, whether justified or not, by at least four, the divider increasing in direct response to the amplification of the praise. On the other hand, many teachers–despite the fact that they are certain they will not be held strictly accountable for all they say–will often stop short of assigning qualities to a particular student that under the broadest conventions of latitude cannot reasonably be attributed to him. The counterplay is for the reviewer, whether rightly or wrongly, to be more keenly sensitive to what is not said than to what is said in behalf of the student, and there is perhaps no more certain way to damn a student than by praising him faintly or reservedly. The result is that even when the commendation is genuine, the praise deserved, and the accomplishment real, the reviewers are unable to determine with any certainty which letters are reliable documents and which are formal exercises. And no one is more fretfully cognizant of the innumerable shortcomings of the letter of recommendation than the teacher who rummages frantically through his word-hoard hunting for language that will alert reviewers to the truly superlative merit of a student deserving uncommon attention.

Beyond the letter of recommendation, we encounter a plethora of criteria that admittedly have at best a tenuous relationship to an

individual's demonstrated competence in a particular discipline or to the intellectual habits and attitudes peculiarly suited to the distinct problems and challenges of the discipline. It is at this level that reviewers can, in the name of "well-roundedness" or "the whole man" give prejudice, whim, and personal priorities authority and power. When the preceding methods of differentiation still leave reviewers with more candidates than positions to be filled, distinctions can then be made on the basis of a variety of more or less respectable, but, in terms of the discipline itself, essentially irrelevant differentia. Here decisions may often be prompted by how highly particular reviewers value the "other interests and activities" or the "professional goals" cited by the candidate. To some, a student who plays the oboe, collects coins, weaves rugs, and spends his vacations on archaeological digs might be considered worthy of selection, whereas others might prefer a former running back who writes music, builds dune buggies, and translates Marshall McLuhan's works into Serbo-Croatian (or perhaps English) in his spare time.

The sophisticated candidate, of course, will not prejudice his case by listing activities not generally acknowledged to be estimable. The lower the predictable correspondence between the value attached to an activity by a student and that attached by a reviewer, the greater the possibility that the activity will not be listed. For example, although certain reviewers might endorse the objectives inherent in a particular political activity, the candidate cannot predict response with sufficient accuracy to risk evoking political biases that, in spite of otherwise impressive credentials, may become decisive. Of course, a shrewd prudence may also deprive him of an admissions ticket, but the candidate knows that, on balance, discretion is the better part of wisdom here, and he pays the reviewers in the coin that he assumes they will accept. There can be no doubt that the criteria at this level are generally least able to endow probability judgments with reliability, are, that is, least able to prophesy the nature of subsequent performance in a particular academic discipline. Nevertheless, it is also true that only at this level can information be admitted which is useful to reviewers willing to modify standards in an effort to control the extent to which the systematic inequities built into the academic structure perpetuate themselves. That is, in the interest of social justice, reviewers may elect to use information in this category to compensate for the elitism that is an integral and constitutive component of the academic structure from top to bottom, supplanting one system of inequities for another, at least until the inequities addressed cease to exist in the society at large.

Of all the supplementary information affecting decision making, universities can at this point in history perhaps least afford to eliminate that

which can be used to prevent the conventionally unqualified from being permanently excluded from the benefits that universities, as credentializing agencies, make possible. In fact, universities may regularly have to invert their hierarchy of selection priorities in order to broaden the base of representation in the student population, endowing information that is of low-order significance (considered in terms of the needs and requirements of particular disciplines) with high-order value.[5] However, when universities neglect certain academic requirements to admit a few young people whom the general society has deprived of opportunities to obtain portfolios insuring admission under "normal" procedures, they not only address a problem created by society at large but also promote inadvertently and undesignedly the perpetuation of exclusivistic practices; although some are admitted, most are still excluded, but the evidence of inclusion takes some of the force out of the charge of exclusion. Token reform, by its very nature, guarantees the survival of restrictive structures, which continue to carry their formal imperatives to logical perfection in the general society.

Although selection committees might insist that no single criterion determines admissibility and that final decisions reflect a painstaking appraisal of the accumulated testimony of a variety of witnesses to intellectual potential, the fact is that, even when the process of adjudication is not complicated by social inequities, none of the witnesses, however ordered or valued, is adequate to the task of predicting with reasonable accuracy how a student may be expected to meet the challenges of a particular subject matter in the future.

Any discussion of categories of selection criteria would be incomplete without a reference to standardized national exams (the Graduate Record Exam, for example), which have virtually usurped the role formerly played by grades and grade-point averages. A student not meeting the preestablished cut-off scores of a specific school will often be automatically eliminated from consideration. Later in the process of selection, when all else fails, reviewers may again resort to the test, giving the nod of approval to candidates according to their rank within the acceptable range of scores. The tests, of course, have much to recommend them. In the first place, they provide quantitative scores on nationally administered exams. Ostensibly, all candidates from all schools sit down to take the battery of tests as equals. At least they do not carry the reputations of their schools

[5] Such an inversion would have no bearing on the proposal presented in chapter 6, since the evaluation system suggested would work effectively, regardless of admission policies.

or professors into the exam with them. Also, the exams are numbered, and, consequently, the readers of the exams are not burdened with personal knowledge about the examinees. Thus, the inherent elitism in other categories of information is obviated, and all presumably sink or swim as ability decides. Moreover, since results are quantified, reviewers can quickly determine acceptable scores and expedite the business of differentiating among differentiable candidates. The lure of established quantities is too attractive to deny; when reference can be made to what appears to be palpable data supplied by candidates and not by external prejudice or idiosyncratic values, the reviewer can exempt himself from responsibility and claim that the candidate, not he, determined admissibility.

Still, in spite of the impressive merits of such tests, they too are unable to provide the basis for meaningful probability judgments. The results of both the general and specific aspects of national tests often tell us only how a particular candidate scored on a particular test on a particular day. That many students are test-wise and course-foolish is a fact as generally acknowledged as it is neglected. And as national tests become increasingly important, the impulse to study tests more and more regularly overwhelms the impulse to acquire knowledge. The rule here as elsewhere is that you learn by doing, and, within certain limits, it appears to be the case that familiarity with tests breeds test competence. There is nothing particularly startling in these revelations; what is startling is the extent to which these tests have achieved hegemonic power in the chambers of selection committees.

At the very least, no test should be considered significant until the reviewers have themselves taken it and evaluated the kinds of achievement or the powers of intellect it is designed to reveal. For a test, as surely as a political pamphlet, embodies a selection of material, a determinate system of emphases, and, consequently, a particular set of distinct meanings. The reviewer must make some effort to delineate the principles and values generating the test as a whole and the displacement and importance of its parts.

Ideally, the reviewer should have access to the judgments determining specific answers, since it is often the case that the wisdom of error is more sophisticated than the wisdom of correct response, and a total of correct responses simply cannot disclose the processes upon which response is based. Anyone who has worked in his leisure on analogy problems, for example, knows not only that he improves in ability to solve them quickly the longer he works at them, but also that in time he moderates the luxuriance of his imagination by restricting the possibilities of appropriate or possible correlation; that is, he learns to trust his immediate associations

and to reduce the number of unusual and freakish possibilities of combination cavorting wantonly in his mind. The point is that a complex mind is frequently animated to intellectual industry by the desire to locate the *concors* in *discordia*, to find, as Samuel Johnson said of metaphysical poets, ways to yoke heterogeneous ideas by violence together. The results of such intellectual sportiveness are often gorgeously ludicrous to someone observing only the results. On the other hand, the pedestrian mind frequently attaches itself to the immediate suggestion, conceives of few possible correlations, and opts for short-range efficiency. But even the student who does not normally formulate unusual hypotheses or find unusual similarities is tempted to try the uncommon when he suspects that those responsible for constructing the test are supersubtle traffickers in the arcane and hopelessly recondite. Paranoia is clearly a natural product of test situations,[6] and to the suspicious mind every question, regardless of difficulty, appears to be loaded.[7] Without any access to the sense informing answers, I do not see how we can submit to being influenced by the results of national tests, no matter how attractively they are tricked out in quantitative garb.

Most national tests now include two or more essay questions, which presumably provide students with opportunities to disclose both their knowledge and their ability to organize, express, and unify ideas. In fact, these questions implicitly place a premium upon the quick recall of specific data and the blurring of fine distinctions. The student who can at will unleash a horde of banal truisms or fire off a few salvos of literary chestnuts will invariably achieve higher points than the student whose extensive knowledge and analytical habits of mind preclude easy generalization and the hasty deployment of arguments requiring substantiation

---

[6] Anxiety is, of course, intensified when the tests are national and can be taken only once or twice.

[7] John Holt relates an interesting anecdote relevant to the issue of student performance and student suspicion: "I am reminded of a second grader I once knew, a bright, troubled, rebellious boy, furiously angry with his parents for reasons I didn't know. I was trying against his will, and therefore unsuccessfully, to teach him to read. One day our school psychologist, a sensitive and sensible woman, gave him a Stanford-Binet intelligence test. Not long afterward we were discussing the boy. She said, 'You know something interesting about his Binet? He got many more questions right at the highest level of the test than he did at the easier levels.' After more thought and talk, we tentatively decided that this boy was afraid to give obvious answers to easy questions for fear that the testers might be playing a trick on him. I have had the same feeling myself, thinking of some seemingly simple question, 'It can't be this easy or they wouldn't have asked it.' " *The Under-Achieving School* (New York: Delta, 1970), p. 68. Published originally by Pitman, 1969.

and qualification. By the preceding, I certainly do not mean to suggest that genuine merit can be determined not by how high, but by how low a student scores on national tests, that quality of mind is inversely related to score. Obviously, bright students regularly achieve high points on such tests; my arguments are exaggerated, on the side of truth I think, in order to remind us that the tests fail to allow for the display of sufficient varieties of intellectual excellence and that many bright students do poorly on such tests. Furthermore, I think that there is something desperately wrong with a system of selection that emphasizes or places confidence in tests which are taken under conditions of high pressure, in a highly competitive and nakedly hostile environment, on one day, and within rigid time limits, and which are constitutionally unfit to reveal the reasons why students performed as they did. In spite of determinate scores, the significance of the tests—which may determine so much about one's future—is largely indeterminate, because too many unknown variables may affect the results. In short, the tests do not test; they only aspire to be tests of achievement and native intellectual ability (whatever that is), and respectable probability judgments cannot be made on the basis of their results.

Probability Judgments and Class Tendencies

3

In the final analysis, all probability judgments are guesses, efforts to anticipate the unknown future from the perspective of a largely uncertain past. The most we can do as brittle beings forced to make decisions in the midst of a radically conditional existence is to base our judgments relating to future performance on a detailed knowledge of past performance, the guiding assumption being that the more narrowly we define the class of past behavior, the more confidently we can anticipate class stability in the future. We imply here a psychical analogue to the physical principle of inertia, and the rule of faith is that a person will act in the future in a manner consistent with what is implicated in the pattern of actions established in the past or fulfill in time the imperatives implicated in antecedent performance, unless either by choice or by chance the ongoing integrity of the class tendencies is interfered with. Prediction, consequently, is always tentative, but without functional type conceptions (and all predictions presuppose type concepts) we could make no judgments at all.[8]

[8] Indeed, the more we learn about linguistic competence and the neurological selectivity of the body, the more apparent it is that we could neither speak nor perceive without type structures. On this matter, Frank Smith has written: ". . .all perception might be regarded as a process of decision making based on significant

Now it seems to me that we have no more reliable source of information on which to base prognostications about subsequent academic behavior than the established pattern of achievement that a student makes manifest in the process of meeting the challenges of specific academic problems over an extended period of time. In other words, only when we know a great deal about the nature and quality of antecedent performance can we begin to narrow the class of action that typifies an individual's intellectual way of being in the world and that imposes probable limits on the range of possible actions in the future. We obviously need more information than we can conceivably gather. For example, it would be useful to know exactly why a student doing well in a variety of subject matters suddenly fails a course not, on the face of it, beyond his previously demonstrated powers; why a student tends to do well in one discipline and not another; or why a student whose marginal competence is indicated by a succession of instructors does exceptionally excellent work in a particular course. Unfortunately, much of what we need to know is inaccessible. We can, however, acquire enough information to make our judgments reasonably reliable.

By this circuitous route we return to grades, the time-honored, traditional manner of typing students, of signalizing the nature of intellectual achievement, each grade a title aspiring (or claiming) to inform the content of academic performance in time with significance and each title contributing to the title of titles, the grade-point average, which, as a summation, acts as the typological fulfillment of antecedent types and,

---

differences detected in the environment. The actual nature of the significant differences is not determined by the physical events themselves, but by the perceiver's rules for distinguishing those events that he wishes to treat as functionally equivalent from those that he wishes to treat as different. . . . In other words, the perceiver brings a highly structured knowledge of the world into every perceptual situation. Rather than saying that he discovers order and regularity that are properties of the environment, it is more appropriate to say that the perceiver imposes his own organization upon the information that reaches his receptor systems. The organization of his knowledge of the world lies in the structure of his cognitive categories and the manner in which they are related. . . ." *Understanding Reading: A Psycholinguistic Analysis of Reading and Learning to Read* (New York: Holt, Rinehart and Winston, 1971), p. 187. Jerome Bruner puts the case succinctly: "We know now, for example, that the nervous system is not the one-way street we thought it was—carrying messages from the environment to the brain, there to be organized into representations of the world. Rather, the brain has a program that is its own, and monitoring orders are sent out from the brain to the sense organs and relay stations specifying priorities for different kinds of environmental messages. Selectivity is the rule and a nervous system, in Lord Ardian's phrase, is as much an editorial hierarchy as it is a system for carrying signals." *On Knowing: Essays for the Left Hand* (New York: Atheneum, 1968), p. 6. Published originally by Harvard University Press, 1962.

furthermore, as the basis for determining how the past can be expected to insinuate itself into the future. In brief, grades are the signs we use to signify past and to imply future performance. They are the chief means by which instructors make what is most important to reliable judgments–the nature of successive performance–available to scrutiny. However, even if instructors were solely concerned with academic performance, they could not possibly rely on grades to represent that performance, since grades can signify nothing distinctly. The basis of meaning is duplication or approximate duplication, that is, one person must duplicate[9] in his mind a meaning generated by another mind. When a sign is used as a shorthand account of something complex, someone must recognize what complexity is contained and what not contained in the account. The complexity isolated by A or 90 is insufficiently definitive; thus understanding, if it occurs at all, is accidental–it is in no way promoted by the sign.

Before this mode of argument is pursued any further, it might be helpful to recapitulate and extend earlier remarks. Grades and grade-point averages are, among other things, probability indicators (at least implicity). Selection committees use transcripts in trying to determine what tendencies in the past will probably be extended into the future. To the individual student, each evaluation notifies him that if he does not take steps to subvert, modify, or improve the tendencies implicated in his past efforts, he will probably reproduce subsequently the levels of performance that he has achieved antecedently. Simply stated, a probability judgment is nothing more than an attempt to subsume the unknown under some known class, on the assumption that the unknown is adumbrated in the known. For our purposes, the principal point is that grades do not make the known or knowable (i.e., specific academic achievement) available to those not assigning the grades; sharability of the knowable is not facilitated by the sign representing the knowable–the grade. The distance between the sign of performance and the nature or content of performance is impassable. And worth repeating is the point that as we move further and further away from the detailed specification of the nature and quality of an individual's specific performance, we move further and further away from evidence on which reasonable probability judgments can be made by students or by others. (No rancor is perhaps more justified than that which a student feels when, after laboring over an essay, he

[9] Perhaps replicate would be more accurate than duplicate, since, to my mind at least, a replica (any very close copy or reproduction) allows for the "more or less," an essential feature of communication; guessing, approximating, paraphrasing are the tools by which we approach another's meaning.

receives it back embellished with "C, this is not satisfactory work"; no information is conveyed, and the student cannot possibly hope to make behavior consonant with professorial expectations.)

Essential to all probability judgments is uniformity or stability of class. Now, whereas chairs, as a class, exhibit considerable class conformity in a variety of manifestations, A's simply do not. The end of a chair is to be conformable to the human body in a very restricted variety of possible configurations, and sharability of meaning is assured by the persistence of traits and the uniformity of opinion concerning the class. If the word "chair" were not adequate to the demands of a particular situation, then the class could be conveniently narrowed to allow for the sharability of precise meanings. Thus, from chair I could proceed to Danish modern chair or to Selby chair (indicating the product of a specific artisan), continually refining specific content by narrowing the class.[10]

Now, we might say that the end of an A is to be conformable to the human mind in a restricted variety of possible mental configurations. The fact is, however, that, unlike the sign "chair," the sign "A" is nonfunctional or insignificant, since there exists no prevailing consensus about the varieties of admirable or desirable mental postures or the distinctive traits characterizing the A class, and no series of adjustments within the sign system will serve to narrow the class of excellence achieved by an individual. Since the meaning of the general class is indeterminate, modifications of the class will be similarly indeterminate. Hence, plus or minus signs cannot do to "A" what Danish modern and Selby can do to "chair."

Moreover, even if opinions about the varieties of excellence inherent in the sign "A" were uniform (and hence sharable), the reader of the sign could have no knowledge of the type of excellence that was actualized (or realized) by an individual student. The general class of A could only be sufficiently narrowed to allow for sensible judgments, if there were generally accepted signs for every possible performance of more or less excellence within the class.[11] (Then, of course, the only remaining task would be to find students who would do work of a quality—and only of a quality—contained within the legitimate horizon of the specified—or understood—subclasses.)

Evaluation, of course, demands the specification of type, since it is only

---

[10] Of course, whenever the class is narrowed, the number of similar traits is increased. For example, a particular chair is like all other chairs in very few specific respects, but it is like all other chairs *of its type* in all (or in a great many) respects.

[11] For example, A++, A+, A+–, A, A –+, A– A––. Of course, to be meaningful, the generally accepted signs would have to have generally accepted values.

when we can know in detail what achievement possibilities have been actualized by a student that we can hope to discover those patterns (or types) of intellectual action characteristic of an individual; without any means of narrowing the type of characteristic action, we can form no reasonable expectations about the nature of future action.

Again, grades, of whatever denomination, do not have self-identical meaning; that is, no grade necessarily means the same thing to two people.[12] The meaning of the grade does not remain constant in the exchange from mind to mind. Two people may, in fact, intend the same meaning by the sign "A," but by itself the sign cannot serve as the indicator of identity or similarity of meaning. What generally happens, of course, is that the reader of the sign assumes that the user of the sign means by it what he would mean if he used the sign; he translates it into his own value system and presumes that meaning is shared, when, in fact, the only thing remaining constant is in the reader's consciousness; he understands the sign "A" to mean what it means to him when he uses it. These are important matters and they deserve our serious attention. Crucially important is the fact that, even if the sign "A" had a *determinate range* of possible meanings (which is simply not the case), the instructor's primary task is to delineate the nature of specific performance within that range.

The problems can be discussed, by way of analogy, in terms of Ferdinand de Saussure's distinction between *langue* and *parole*,[13] langue being a particular language (say, English) with its lexicon and syntactical rules governing arrangement possibilities, etc., and parole being a specific utterance actualizing a selected few of the language possibilities. Not until words are selected and arranged in an effort to embody a specific intention or meaning can we have an utterance of determinate meaning. That is, a specific statement has to be made before an intended verbal meaning can be discovered or before its "intention" can be known by others.[14]

---

[12] Essential to self-identity, as the term is used here, is persistence through time and consciousness of determinate meaning; that is, the meaning is identical with itself from one moment (one time) to another and, inferentially, from one mind to another.

[13] As expressed in *Cours de linguistique générale*, the book which a group of the linguist's students constructed from lecture notes and published in 1915. Translated by Wade Baskin and published by McGraw-Hill in 1966.

[14] The reader should note that throughout I am insisting on the "logical" priority of intention (whole meaning) in statements. That is, intention determines the selection and arrangement of words (and, hence, is logically prior to them) and is what is realized by the words expressed. With regard to the reader, I am insisting that his understanding of the parts of a statement, in their regular unfolding, depends upon a conception of the type of whole meaning to which the parts belong; otherwise, the

Once a statement is made–"I am going to the store," for example–an infinite number of linguistic possibilities is reduced to a particular set of combinations; a specific choice of words is made, and a particular arrangement is selected.[15] (Radically different possibilities would have been actualized, for instance, if I had said: "*Am I* going to the store?" or "*You* are going to *a* store.") Even though I have severely limited meaning possibilities by saying that "I am going to the store," I cannot depend on these words alone to convey a determinate meaning. In fact, a reader who has access only to these particular words cannot possibly know what those words mean, for they clearly do not mean all that they could possibly mean. The point is that a statement as ostensibly innocuous as "I am going to the store" is, by itself, indeterminate, capable of supporting a rich multiplicity of hypothetical meanings. The point is easily demonstrated; say the sentence six times, emphasizing a different word with each repetition. Is the statement a matter of fact declaration? an act of defiance? In isolation, the statement cannot disclose whether it is spoken by a man informing his wife that he is going to the store, *not* to the tavern, or that he, *not* someone else (contrary to assumption or wish), is going to the store. An auditor would understand the remark better than a reader, since he would have access to tone, emphasis, inflection, and perhaps other determinants of meaning. Similarly, a person who had been present at many verbal interchanges between two people would be in a highly favorable position to explain what was particularly meant when one said "I am going to the store" to the other at this particular moment. In brief, meaning depends upon familiarity with something more than the meaning of individual words in a particular arrangement. We must be able to narrow the class to which the statement belongs, and we can only do this when we know not what the words can mean, but what they probably mean. The class of probable meanings is narrowed when we can determine the whole

---

reader would not be able to understand the individual words as parts of a meaning (or any meaning in particular). For both writer and reader (speaker and listener) a type of whole meaning must be construed before the meaning of the words can be known. From very meager beginnings (on the basis of very few verbal facts) we intuit the type of whole meaning that demands fulfillment in time.

[15] Throughout this section, I owe some debt to E. D. Hirsch's *Validity in Interpretation* (Yale University Press, 1967), which I had read many months before I even thought of writing this book. In reviewing his book for a purpose totally unrelated to this piece on evaluation, I discovered that I had unconsciously taken as my model sentence one embarrassingly similar to his, "I am going to town today." The similarity is quite accidental, and I, of course, am fully responsible for the uses to which I put the model sentence. The reader who detects any faults in my argument should not assume that he has, similarly, found weaknesses in Hirsch's case.

(or the type of whole) of which the statement is a part. Consequently, the more information we have about the conditions generating the statement, the more certainly we can infer the probable meaning of the statement. In the present case, we can only begin to speak with authority about the meaning of "I am going to the store," when we know, for example, who (i.e., what sort of person) says the words to whom, under what conditions, to accomplish what ends. For our purposes, the point worth remembering is that *langue* is the system of conventions (relatively stable, but changing in time) determining the possibilities of combination, whereas *parole* is the actualization of possibilities in a determinate meaning. The task now is to relate the preceding discussion to the problem of grades, specifying the analogical counterparts to langue and parole.

We begin with the assumption that appears to underlie the whole grading system, namely, that academic performance implies a convention system, implies, that is, types of knowable performances; skills, knowledge, and intellectual powers can be combined in many diverse ways to produce, finally, actual performances falling within a range of performance possibilities for a specific task, discipline, course, etc., and identifiable by the letters from A to F. The range of performance possibilities intrinsic to academic achievement is not the same as the range intrinsic to achievement in, let us say, fishing. Briefly, then, we make "sentences" out of the "language" of academic performance possibilities by writing essays, exams, and so forth. The actual performance of an individual student is the *parole*, but although the student can make a variety of performance statements, the instructor can only make, for example, five statements about the many performances (i.e., A, B, C, D, or F). Essentially, then, A, B, C, D, and F are quasi-paroles (i.e., shorthand "statements" about particular, determinate performances[16]). Nevertheless, saying "A" is very much like saying "I am going to the store." An A is as radically incompetent to designate the nature of actual performance as "I am going to the store" is incompetent to convey determinate meaning. The conscientious instructor who listens patiently in the quiet and lonely isolation of his office to the seemingly interminable deliberations of the parliament of his mind and then rushes, desperately and with some misgivings, to a C is accomplishing little more than would be accomplished by a monkey which, in the process

[16] In an effort to simplify matters, I have chosen not to examine in detail the full implications of the distinctions that I have made, since my focus is not on complex, linguistic categories, but on grades as illuminated by an analogy to *langue* and *parole*. Grades are interpreted as elliptical sentences and as signs of complex, actual performances; consequently, they can be discussed as meaning units comparable to sentences like "I am going to the store."

of randomly pecking typewriter keys, accidentally produced a distinguishable English sentence. The accidental sentence would be deprived of a definitive context[17] and, hence, of determinate meaning; from the evidence of the sentence, no principle for excluding or including possible meanings could be discovered or inferred. As far as the instructor is concerned, it is not that he does not intend to indicate the nature of specific performance; he simply cannot do it with the sign "C."

Each letter grade is obliged to represent a wide range of performance possibilities. And even though a particular grade is meant, in the mind of the instructor, to designate the actualization of particular possibilities, the reader of that sign has no way of knowing what precisely is being designated, and this is true even if the reader happens to share the instructor's views on C possibilities. An infinite variety of sentences can be constructed from the English language, and although each sentence is by itself mysterious, we can come to know what each most probably means. We do, after all, manage to understand a good deal of what we say to each other. Undoubtedly, there are fewer academic performance possibilities (within specific disciplines and assignments) than sentence possibilities, but the fact is that the numerous performance possibilities cannot possibly be delineated by the four or five (or seven or eight) signs that are used to signalize performance. A sentence normally does not stand by itself; we can usually come to know the verbal and extraverbal context informing the specific sentence with specific powers. On the other hand, a C, as recorded on a grade sheet, has no context; we, as readers of the sign, have no access to information that would enable us to view a particular C as a specific type of performance. C, in short, is contextually free and, consequently, essentially meaningless, at least in the sense that C conveys no knowable range of probable meanings. In practice, the reader—as stated earlier—applies a meaning to the sign that squares with his personal conception of the performance possibilities falling within the range of the sign; the inescapable fact is that the reader more than half creates the meaning for the C that he observes.

Essentially, grades are typing errors, that is, grades attempt to specify types of academic behavior, but they are totally inadequate to the task, since they do not resonate with specific type meaning. The type cannot be

[17] What is true of this statement is also true of any isolated statement. By "context" I mean a good deal more than the sentences immediately preceding and following the statement under discussion and more than, say, the particular "setting" of the statement. Context is a type of whole meaning, in relation to which particular statements are necessary parts. Context is, then, a constitutive, formal principle of meaning.

known by the agent of typification. In slightly different—and perhaps more respectable—terms, grades can be classified as category mistakes, futile efforts to denote determinate behavior by means of indeterminate signs. And as soon as mathematical values (i.e., specific normative weights) are applied to these signs, we call upon our heads the charge of misplaced concreteness; we are guilty of taking the flesh for the spirit, of assuming that real qualitative differences achieve ontological status only when they acquire quantitative palpability.

4

Confusion Compounded

In the preceding chapter, I was willing to assume, for the sake of argument, that there is universal agreement about the diverse types of behavior that can legitimately be typed A, B, C, D, or F. The hypothesis was adopted in order to demonstrate that even under optimum conditions grades prove to be fundamentally incapable of performing their assigned duty, principally because they cannot narrow the class of behavior with sufficient precision to promote reasonable probability judgments. The assumption is, of course, unwarranted; there is no such consensus.

Once we being to look at the various criteria that are invoked in defense of particular grades and at the number of nonacademic, irrational, quixotic, or idiosyncratic factors that play a role in their selection, we immediately recognize that the apparently solid edifice of grades is supported by an unruly mob of partisans united only in an effort to keep the inadequate structure from toppling. From a distance, the cacophonous roar of conflicting opinions sounds like the he-bellowing voice of one man crying, "Maintain the deception!" It would be impossible to compile an exhaustive list of all the factors affecting grades, but, before going on to other matters, a few should be mentioned.

By now most of us have either read about or been involved in tests which require a group of teachers to read several essays and to rank them

by grade or value. Invariably, there is considerable disagreement among the teachers about the best and worst essays, as well as about the essays in between. When the different judgments are discussed at length, few alter their views, and unanimity of opinion is seldom achieved. Among other things, these tests indicate that teachers bring radically different predispositions and standards to the examination of student essays and that unchanging material can be interpreted in marvelously diverse ways and can be obliged to conform to widely varying criteria of value.

One teacher, for example, may have a comma splice quota; beyond the quota, failure opens its gaping maw. Another may stress tight, logical structure. Still another might reward Latinate diction and polysyllabic bombast, and another might value a no-nonsense, hard-hitting, argumentative truculence. And on and on. The tests are given, discussed, and forgotten. Each teacher learns something from the experience, perhaps that he ought not to discuss his criteria so candidly in the future or perhaps that his colleagues are what he thought them to be—corruptors of youth and contributors to the decay of academic standards.

At times, these tests promote an evanescent interest in grading reform, and some perfunctory attempt is made to standardize criteria or to construct objective exams. But rarely can a committee agree on specific criteria, and general criteria, even when widely accepted, are practically useless, since excellence is multifaceted and is achieved by diverse methods. We have to know what students have actually accomplished and how the general criteria have been realized in specific cases.

"Objective" tests, on the other hand, are seldom useful to instructors who take their discipline seriously, since they convey so little information of the sort that instructors are most interested in having. That is, it is clear that some wrong answers are less significant than others; an error must be traced to its sources before it can be properly evaluated. Answers are wrong for different reasons, and errors are differently sophisticated. What the instructor most wants to know, I think, is how well a student is able to think as a mathematician, historian, or whatever, and whether he can generate and resolve respectable intellectual problems. Among other things, he should know whether the error was careless (a simple mistake in addition at some point in a complex mathematical computation), whether it was due to a temporary loss of memory (a fact or formula could have been applied if it were recalled; the intellectual operations could have been successfully performed if the exam, say, were open-book), or whether the problem was beyond the intellectual powers of the student (even with all the basic information supplied, he would have been unable to work it out).

On the whole, established procedures are not modified, and grades

continue to be informed by invisible or dubiously respectable factors. Some instructors adjust final grades to square with performance at the end of the course, that is, they tip the balance in the direction of improvement, weighing work at the end of the term more heavily than early work. Again, the importance of participation in class discussions varies from instructor to instructor, and grades vary accordingly. Also, a grade on a particular essay may depend upon its location in the pile. An instructor who has just read three exceptionally bad papers is likely to greet the mediocre paper which follows with enthusiastic favor. (The fact is that an instructor is intellectually ready to grade papers just before he finishes reading them; by the time he reaches the end of the pile, he is prepared to bring to the reading of the papers a fairly clear sense of what students may reasonably be expected to do with the assignment. At the end, he has established a functional and useful set of probability expectations.)

Another complication is introduced by instructors who, while proclaiming support of intellectual pluralism, tend to bestow favor only on that work which confirms or adds support to their own pet theories. Also, instructors generally tend to underestimate the extent to which students are able to work out strategies that are guaranteed to win approval. Grades, then, often reflect a student's ability to pick up the professorial cues to success. Consequently, a dull, plodding paper decorated with gems inadvertently supplied by the instructor may get higher marks than one which carries to term a complex thesis incompatible with the views of the instructor. The holders of Phi Beta Kappa keys are often little more than clever tacticians with a facility for reading the chief plays of the instructor's game plan. When the end of the enterprise is a high grade, the wisest policy is to abandon novel, risky, dangerous, and, hence, grade-point-average-threatening ideas. If an instructor happens to believe, for instance, that *Othello* no more exemplifies the theme of love versus reason than punching someone in the nose exemplifies the theme of war, then no argument supporting the rejected thesis, however subtly it is managed, is likely to win instructorial favor. Transcripts are strewn with low grades that have their origin in intellectual temerity. But there is no end to the job of enumerating the nonrational or idiosyncratic factors which make their influence felt in the grading process, and I am sure that every reader can immediately add many to the foregoing list. The point again is simply that a grade cannot by itself disclose the rationale determining its selection; when the rationale can be presented, the grade is superfluous.

In spite of all that can be said about the absurdities inherent in the grading system, the belief persists that, on the whole, it works. It is assumed that over the long haul of four years or so the student

consistently rewarded with A's is a good student, one who will probably continue to do good work on the graduate level. Of course, the assumption is largely self-proving, since those who have "good" credentials go on to graduate school. But the idea is that a distinct type of academic behavior is clearly delineated by the grade-point average; some kind of excellence presumably persists through the process of being judged in a variety of disciplines by various minds with diverse standards. The grade-point average amounts to a kind of indicator of true academic worth; if, after being subjected to such diversity, a student emerges with a high average, then it is assumed that intellectual excellence is the uniformity amidst the diversity. Undermining this logic is the fact that no single subscore of the total score can be accepted as trustworthy; furthermore, there is no way of differentiating among the subscores in terms of more or less reliability or validity. Since we cannot know what standards determined what grades, we obviously cannot establish a rational basis for discriminating among kinds of grades.

Without knowing what kinds of criteria are being employed to distinguish what from what, educators have consented to believe in the discriminative and predictive powers of grades. Consequently, many bright and competent students of generally unknown capabilities have regularly advanced on the academic ladder, and, more important, many others, also bright and competent, have been deprived of academic mobility altogether. The alarming fact is that grades continue to be used, even though they cannot adequately typify behavior or make evident the bases of success or failure.

As I stated at the outset, the inadequacies of grades are greatly aggravated when they rise generally throughout the nation. The higher they rise, the less significant they become (and, paradoxically, the more they are needed). As a result, decisions concerning merit are based on evidence further and further removed from the nature of academic performance as manifested in a succession of intellectual efforts. The report concerning grades at San Francisco State, properly considered, is an invitation to administrators and teachers to reassess evaluation procedures. It should provoke serious debate about the value and efficacy of grades and encourage the formulation of proposals for the radical revision of established evaluation procedures, not instigate face-saving attempts to explain away the local occurrence of a national phenomenon by parading upper-division and bright students before us. We only solve the problems that we address, and we do not address the problem of grades by only asking why they are higher now than in the past. High grades can be explained in various ways—some of which have been outlined above—but

whether they are high or low, the fact is that they are radically unsuited to do their assigned work. Only when this is frankly admitted are we in a position to begin constructive reform. Moreover, it should be clear by now that since grades are endemic to education, no reform can be useful unless it is generally adopted.

5

Alternatives to Grades

Among the intramural (and not widely publicized) demands that students across the nation are making of universities is one calling for a radical revision of the traditional means of defining achievement in courses (students sometimes insisting that grades be eliminated entirely), and many schools, aware of the growing dissatisfaction with grades, have recently expressed a willingness to consider—and sometimes adopt—innovative methods of evaluating students.

Except in a handful of cases, however, the "new" systems are little more than ill-disguised variations of the grading system, and virtually all the innovations suffer from the same shortcomings that afflict grades. For example, a few years ago a prestigious university, responding to demands for reform, instituted a phrase system. For A, B, C, D, and F, the school substituted a graduated series of terms to indicate levels of achievement. Henceforth, students completed courses "with high distinction," "with distinction," "with little distinction," "with no distinction," "with distinctly no distinction," and so on.[18] Of course, "distinction" is not defined, and individual performance is not typified by the specification of

[18] These are not exactly the phrases used, but they are not very far removed from those actually employed.

characteristic levels or kinds of performance. One meaningless set of signs is replaced by another. No matter how thoroughly performance possibilities within each category are defined, we as readers of a specific category cannot know with any confidence what possibilities were realized in what ways by an individual. "With high distinction" discloses the nature of performance no more accurately than an A. Like "I am going to the store," the phrase can support far too many legitimate interpretations, and none of them solidly. Indeed, all the arguments attacking grades can be directed with equal validity against the "distinction" gradations. A specific term ("with little distinction," for instance) is not a knowable subtype of a known whole type. The system is finally an exercise in self-deception; the same characters (albeit with different names) play the same parts in the same inept fashion.

In an effort to refine discriminations, a well-known law school adopted new evaluation procedures. Law students in each class were awarded at the end of the term a T, M, or L, signifying respectively Top, Middle, and Low. Also, by administrative fiat, it was predetermined that 10 percent of the students would be awarded T, 80 percent M, and 10 percent L. That is, the distribution of grades was established before any work was submitted. (Surely this amounts to a form of prior restraint or judgment on the basis of facts not in evidence.) In any class of thirty students, then, three would receive T's, twenty-four M's, and three L's. Animating this system, undoubtedly, is a vicious desire to provoke intense competition among students already frighteningly competitive. With few T's available, all students presumably will strive frenetically to acquire them. But regardless of how assiduously students work, the fact remains that 90 percent of them must inevitably come short of their goal. When charted, however, each class will be represented by the comely undulation of a bell-shaped curve, which by itself compels admiration in many minds, since it displays in a graphically cogent form a fact about distribution generally accepted, especially by those faintly acquainted with statistical problems. I was once one of two readers charged with the task of reading undergraduate papers for a lecturer who read only the papers of the graduate students enrolled in the course. The class was divided alphabetically, and each reader examined half of the papers. When we brought our returns to the lecturer, he asked us both to make graphs of our grades. As matters turned out, my scores were exceedingly reluctant to assume bell-shapedness and were, of course, immediately suspect. In defense of what the lecturer clearly understood to be the indefensible, I said that although the students had been divided alphabetically, there was no prima facie reason to assume that talent had similarly been distributed along alphabetical lines and that the

grades in an elected, specialized course for seniors would not necessarily be distributed in the proportions that might conceivably be expected in a required, general course for freshmen. From the glare of the lecturer, I turned to the basilisk for comfort. My feeble arguments were dismissed as they were articulated, and the lecturer quickly determined that reasoned rebuttal would be wasted on me, since I was patently one upon whom the accumulated evidence of statistical research had made no impact. The moral of this homely tale is that evidence is frequently confronted as evidence only when it appears in the forms constituted by our wishes or expectations and, I suppose, that no expectations are more unshakable than pseudoscientific ones.

Even without the preordained distribution figures, the T M L system would be a failure, if only because it cannot differentiate by any sign among students in the middle, the category allowing the greatest range of performance possibilities. But what is true of the middle is true also of the other categories, since even in the Top category, some students are clearly "topper" than others and in ways which cannot be expressed by the letter "T." Three signs (chosen to represent the actualization of performance possibilities) are asked to do what was formerly done incompetently by five (i.e., A,B,C,D,F). Once again, T M L fails for precisely the same reasons that A,B,C,D,F fails: each categorical statement supports too many viable, hypothetical meanings, and even if we were to admit that three of thirty students could be expected to do conspicuously excellent work, the fact remains that the "T" sign simply cannot tell us what kinds of excellence were made conspicuous, how different low "T" excellence was from, say, high "M" mediocrity, or what criteria had determined what was going to be called excellence. Moreover, even if T's were awarded to those students, and only those students, who "wrote in clear and forceful prose persuasive briefs supported by the most important and relevant precedents, etc.," we still would not know what we most needed to know, that is, what intellectual habits and powers combined in what ways to select and arrange what material to accomplish what specific ends of what particular value. Clearly, the T M L system obscures rather than illuminates meaningful distinctions among differentiable achievements.

The reform most actively supported by many students is the pass-fail system, which is self-explanatory—at the end of the term, students either pass or fail the course. The schools which have adopted this system have usually done so only in a piecemeal or halfhearted fashion; that is, most schools do not offer all courses on a pass-fail basis. For the most part and at most institutions, students are allowed a specified number of pass-fail courses a year, and they generally can take only nonmajor courses

pass-fail. For example, a mathematics major may take an English course, but not a mathematics (or related) course, pass-fail; similarly, an English major may take a mathematics course, but not a literature course, pass-fail, and so on. The potential threat to traditional grading procedures implicit in the pass-fail system is effaced when universities tolerate the reform only within these severely restricted limits. Universities are saying, in effect, that they will institutionalize the reform only if it is rendered harmless. What this means, of course, is that the school strengthens an inadequate system (grades) while appearing to undermine it. Dissent is appeased, and the old grading system, which provoked the demand for change in the first place, exerts its influence with rejuvenated vigor. Grades are more, not less important than they were before, since—so the argument runs—pass-fail serves as a kind of alchemical agent separating the dross from the gold; that is, with pass-fail allowances, grade-point averages are no longer adversely affected by as many nonmajor grades as they were in the past. Also, because students can presumably devote additional time and energy to work in the major when they are required to obtain only a passing grade in another course, "major" grades may be considered as accurately reflecting true levels of academic competence. Competition for high marks is often fortified, not weakened, by pass-fail.

Even so, the system is actively supported and promoted by many students, because they can derive from it several real and substantial benefits. In spite of the covert inducement, many students do not take pass-fail courses so that they can devote more time to their majors. Instead, they often elect to take courses which they normally would be unwilling to try if their achievement in them would finally become part of their grade-point averages. That is, a chemistry major may try an upper-division Shakespeare course if he knows that the market value of his academic record will not be decreased by his performance in the course. In general, pass-fail courses tend to relax tensions; the classroom timidity that competition for grades encourages is diminished in pass-fail courses, and students are frequently emboldened to ask tough and relevant questions, to risk saying that they do not understand what is being presented, and to behave generally like people whose primary objective is to learn. Again, students can afford to work out the implications of novel, risky, and normally dangerous hypotheses in their essays, when they can reasonably assume that however much an instructor may dislike the nature or tendencies of their views, he will not fail outright those who have worked diligently and completed all the required work. In short, students stop playing education safe when the risks are not prohibitively great; pusillanimity is diminished in proportion to the reduction in fear.

At best, however, the pass-fail courses allow but a brief respite from frantic competition for grades. And whatever may motivate a student to take courses pass-fail, he cannot escape the fact that academic advancement or occupational placement is still dependent upon his grades, not upon his performance in elected courses. As currently adopted by most universities, pass-fail simply reinforces the grading procedures it was designed to subvert. In fact, many of the incidental benefits of pass-fail are virtually cancelled when universities, fearing intellectual indolence and the breakdown of standards, assign to pass the value of a C or D, thereby discouraging those students with high cumulative scores from taking courses on a pass-fail basis altogether. The fact is that universities in general have a low toleration threshold for forms of evaluation latently inimical to grades.

But apart from its adventitious merits or the ends that universities may make it serve, pass-fail is, as an evaluation device or as a means of denoting achievement, completely worthless, primarily because it conveys no sharable meanings and promotes no sensible guesses about the nature of specific performance, no reasonable inferences that can be legitimized by appealing to even a personal, idiosyncratic hierarchy of performance values ("pass" is attached to excellent and mediocre work alike).

By academic action (i.e., by essays, exams, conversation, etc.) students discriminate among themselves; intellectual habits and powers are made manifest in action, and one of the instructor's most important tasks is to describe and evaluate that action as accurately and fairly as he can. To rank under one rubric (pass) a multiplicity of achievements is, to my mind, to abdicate one's responsibility as a teacher, to filter observable shades of difference through a monochromatic lens, supplying a uniformity that is not found. More important, the impulse to excel, to extend knowledge, to transcend previously established limits, initially encouraged by the liberating tendencies of pass-fail, must inevitably be thwarted when "pass" emerges at the end of all effort as the great leveler. Of the systems of evaluation considered, pass-fail is patently the crudest. It extends rather than limits the number of performance possibilities realizable within its selected terms of distinction. With fewer signs of difference, we can make fewer reasonable (and more unreasonable) inferences about the nature of specific performance and fewer sensible anticipations of the nature of subsequent performance. What cannot be accomplished by five or three signs certainly cannot be accomplished by two.

In noting grading reforms, we have followed the direction of popularity to signs of progressively diminishing precision. But throughout the immediately preceding discussion we have in reality been considering only

more or less subtle variations of the A,B,C,D,F system, and whether we call A's pheasants and F's pigeons, employ a selected few phrases, or bifurcate performance into two possible categories, we are locked into a single system of evaluation. Inhering in each of the variations are typing errors, category mistakes, and examples of misplaced concreteness. As long as behavior is represented by a restricted and unchanging set of signs or terms, evaluation and prediction will be grounded in mystery and uncertainty.

A Proposal: Written Reports

A critic, of course, is not obliged to supply remedies for the deficiencies that he notices; his first, necessary, immediate, and perhaps only task is to expose weakness, folly, and error. Nevertheless, most critics manage to assert by implication what they may neglect to make explicit. In this essay a specific reform has been implicated in almost everything that has been said in opposition to grades. Indeed, to my mind, the preceding arguments have their proper culmination in a particular proposal, one that has been silently but conspicuously present throughout, namely, an evaluation system based on written reports for each student in each class. These reports would form the substance of the students' academic records and would be the principal resource of reviewers responsible for making probability judgments.

In this concluding section I shall briefly outline what I take to be the nature and merits of the system, noting at the outset that written evaluations can take various forms without violating the integrity of the system itself. Under this system, then, an instructor would be obliged to comment on each piece of work submitted to him by each student in his courses. Since specific intellectual powers, habits, and processes are embodied in the performance of academic assignments, the instructor would be asked to describe as accurately as he could the nature of each

student's work and to relate that performance to whatever standards, values, priorities, and expectations, he considered appropriate.

Presumably, no assignment is purposeless; instructors, it is assumed, make assignments in order to accomplish distinct goals or to provoke certain kinds of activity. As pointed out earlier, every exam (or essay assignment, for that matter) is designed to do some things and not others, and instructors can make the design available to students and subsequent reviewers and can also indicate the ways in which individual achievement satisfies the demands inherent in the assignment. In short, instructors are asked to describe what a student has done and briefly evaluate achievement in the light of the formal imperatives implicit in the assignment and according to articulated criteria of excellence. (Of course, the line separating description and evaluation is not a clear one—at least not necessarily—since each instructor brings to description value priorities which often determine what he sees or at least what he pays attention to and since evaluation inevitably implies delineation of what is being valued. What is crucial is not the rigid separation of categories—desirable perhaps, but impossible to achieve—but the detailed verbal account of the assignment and of the student's performance, for however intimately connected or clearly divided description and evaluation are in any particular report, the general rule here is that the bases of judgment are revealed in direct proportion to the inclusiveness and thoroughness of the report; the more the writer says about the assignment and the performance from the standpoint of whatever priorities and criteria, the more he discloses about both the work and the principles constituting judgment.)

Governed by the understanding that the readers of his report will have some familiarity with its subject matter, the instructor will simply discuss the achievement of the student in the terms and according to the principles with which he is most comfortable. However, if a particular exam is designed to determine simply how well students can recall specific facts, then the report should indicate only that a given student faithfully remembered, say, six of twenty facts, fifteen of twenty facts, etc. When an exam relies heavily upon such questions, the fact should appear in the report, along with whatever explanations of the results or defenses of the exam itself seem appropriate to the instructor. In referring to standardized national exams, I made the point that, although error is often more intellectually respectable than correct response, the exams provide no intrinsic means for distinguishing between sophisticated folly and naive wisdom. Now what is true of national exams is similarly true of local exams of the same kind. But with written reports, instructors clearly can make the necessary distinctions if they pay attention to and attempt to

define the intellectual processes responsible for particular responses.

I once asked a young girl who had recently been working with multiplication tables if she could multiply twenty-five by twenty-five. She had not before encountered a problem quite as complex as this, but after pausing a minute to consider and calculate, she came up with sixty-five. The "correct" answer, of course, was 625, and I could not understand how she had arrived at such a grossly inaccurate figure. When I asked her how much twenty-five plus twenty-five was, she immediately said fifty, without realizing that that figure was only slightly less than her total for twenty-five times twenty-five. Now, if I were simply interested in whether she could answer my question "correctly" or not, I would have discovered nothing about the ingenuity of her mind. The fact is that her answer is perfectly sensible, and her result is absolutely "correct"; as long as the figures are construed as they were by the young girl, the answer will be sixty-five. Her reasoning: five times five is twenty-five; write the five below the line and carry the two; now add the three two's in the left-hand column to get six and, hence, a total of sixty-five. Confronted with a totally new mathematical situation (well, almost totally new), the child rapidly conceived a working hypothesis that would yield a confirmable result; this, I submit, is sophisticated problem solving, but I would have discovered nothing about her genius if I had not paid attention to what she was doing; moreover, once I knew how she had construed the mathematical event, twenty-five times twenty-five, we could march briskly together to that other "correct" answer, 625.

Many exams are constructed to permit the hasty determination of the number of correct answers; by choice the instructor is preoccupied with result, not process. No change in the evaluation process will necessarily induce an instructor to concentrate on intellectual process, but without any genuine evidence concerning the kinds of intellectual talents represented in his classes, the instructor might frankly admit his inability to typify students, to define characteristic intellectual habits and powers, and might note only how students scored on tests requiring the recall of specific data.

In a real sense, the written report only asks that an instructor note how students distinguish themselves from one another in a discipline in which, by training and experience, he is presumably competent. The reform assumes that teachers are adequately prepared to comment on achievement within their disciplines, differentiating among distinct kinds of excellence and deficiency. Furthermore, only by using terms capable of conveying sharable meanings can they hope to insure the self-identity of meaning through time and in a variety of consciousnesses. No amount of

training and experience will endow instructors with powers to convey an unchanging meaning to different minds by means of a restricted set of conventional signs, each fostering a rich multiplicity of more or less legitimate interpretations. (Written reports do not guarantee that precise meanings will be shared; they simply provide the necessary conditions for the sharability of determinate meanings, for the transmission of relatively precise meanings.)

By its nature, the reform cannot prescribe the content of the reports; unlike A,B,C,D,F, written reports can signify achievement only in the terms that achievement and personal intellectual priorities make necessary, that is, grades dictate that achievements fit five preexisting categories, whereas written reports arrive at the unprescribed statements dictated by performance.

Nevertheless, although the content is not predetermined, the value of the reports is directly dependent upon the content—in precisely the same way that the value of any review or critique is dependent upon content. Reports, that is, are valued more highly as they account—accurately or inaccurately, from sound or unsound premises—for more particulars. When little can be said by instructors, little can be known or considered by reviewers of the reports.

The mechanics of the system are easy to describe. The length of the reports would vary, but most reports would not have to exceed two or three hundred words. At the end of the term, the instructor would write a summary report for each student and would submit it, along with comments on particular assignments, to the registrar. In general, each student would have approximately two or three pages of comment for each course in which he was enrolled. The reports would become a part of his academic record, and in place of a computerized grade sheet he would receive copies of the several reports. Ideally, the contents of the reports would not startle the student, since most of what he finds in them will already have been communicated to him during the term at the bottom of his papers, at the end of his exams, or in conferences with his instructors.

In his summary report the instructor would note all those evidences of intellectual competence that could not conveniently be introduced elsewhere, such as the quality of mind displayed in classroom discussions or during office hours. Intellectual powers are variously manifested, and academic excellence, in and out of the classroom, should, of course, be represented in the account of the student's overall performance. The student who regularly sparks discussion, raising challenging questions and suggesting interesting ways of construing events, deserves to have his demonstrated excellence recorded in his official transcript.

Just as, among faculty members, there is no positive correlation between quality of mind and the fact of published work, so there is, among students, no absolute correlation between quality of mind and performance on any given written assignment. Nevertheless, students, like people generally, will exhibit powers of considerable class conformity in a variety of intellectual efforts. Consequently, the summary report should be the most comprehensive account of characteristic intellectual habits and resources (as disclosed in intellectual actions) and of typical levels of performance.

Personality is known chiefly through the persistence in time of determinate traits of behavior; similarly, intellectual character is defined by a succession of endeavors exhibiting determinate kinds of redundancy. By actions we typify ourselves, and the instructor's task, in large part, is to discriminate among types of intellectual performances, noting which traits carry over, continuous and changing, from activity to activity. Of course, instructors make mistakes, and a variety of factors, academic and nonacademic (financial, emotional, etc.), can preclude the realization by students of type tendencies. However, consistency from report to report serves as something of a check against error, as a means of moderating the influence on reviewers of anomalous traits. The recurrence of types over a period of four or five years in a variety of reports written by a wide variety of instructors unfamiliar with the details of one another's evaluations gives to each of the compatible judgments a kind of self-certifying validity—each local judgment resonates with general authority, because each implicates all the others with which it is consonant—and provides a reasonably reliable basis for prediction, for anticipating the ways in which the past will be carried into the future.

A further check against improper type-casting would be provided if students were allowed to include in their records one or more examples of their academic work which, in their opinion, best represented their skills and powers, especially when they were convinced that the instructorial report grossly misrepresented their achievements. (As a general practice, it would perhaps be advisable for reviewers to consider at least one example of a student's work directly, without having it filtered through the terms of an instructor.)

The essential point is that with or without trustworthy information, reviewers are going to make decisions on the basis of type conceptions, and, although written reports do not guarantee the emergence of valid types, they at least make real confusion accessible and provide information infinitely more valuable than grades, national test scores, letters of recommendation, and so forth, since they disclose what criteria of

evaluation are being applied to what material. Rejuvenated interest in academic subjects, emotional problems, financial shortages, new insights, etc., can all subvert established and characteristic tendencies and instigate the development of new performance possibilities, but the detailed accounts of realized academic achievement, for all their inherent shortcomings, are best equipped to handle the complex problems involved in the process of evaluation. Even when no pattern of consistent action is clearly discernible in the student's record, the confusing disarray of discordant, disparate judgments is highly informative, providing reviewers with more useful evidence than could possibly be found in any of the other forms of evaluation discussed in this paper.

II

When in chapter 5, I examined a variety of alternatives to conventional grading practices, I neglected to comment on one alternative which is rapidly gaining widespread support from many students and educational critics, primarily because I thought that, for my purposes, it could best be discussed in relation to written evaluations. The common term for this alternative is "credit/no record." Its nature can be briefly described: at the end of the term, those students who did satisfactory work in the course would be given credit, and those who for one reason or another did not meet minimum requirements would be notified by the instructor that they were not going to receive course credit; only the names of those receiving credit for the course would be sent to the registrar's office.

Of all the proposals currently being suggested as alternatives to traditional letter grades, credit/no record is, in my opinion, clearly the best. Although on the surface it would appear to eliminate evaluation altogether, the proposal—directed primarily at the administrative and "judicial" functions which normative signs serve (decisions which relate to whether the student stays in school, how many courses he may take, whether he shall retain a scholarship, etc., are usually made on the basis of grade record)—strongly emphasizes the educational importance of instructorial evaluations, only noting, by way of qualification, that the instructor's evaluation of student work is essentially a matter of concern to the student and the instructor. Far from discouraging evaluation, credit/no record enjoins the instructor to comment extensively on student work. Starting from where the student is and using his training, skill, intelligence, and experience primarily for diagnostic purposes, the instructor will indicate in his comments how, for example, the student might bring to

excellence what he has already carried a considerable distance. In short, the instructor will continue to do what, by virtue of study, training, and experience, he is presumably best equipped to do, namely, impart what he has learned and discuss student work in terms of its inherent shortcomings and possibilities. On the other side, he will simply omit to do what he does least well, what is demeaning to both him and the students, and what is least relevant to the educational process and his role as an instructor, namely, attach at the end of each term a normative sign to each student in his class, not for the benefit of the student, but for the convenience of administrators. To the sponsors of this reform, grading in any form is irrelevant to education; as an academic enterprise, grading is, for instructor and student alike, degrading. Under this proposal, then, the detailed evaluation (presented orally or in written form) is a fundamental condition of effective teaching and an absolutely essential component of the teaching-learning process.

There is virtually nothing in the credit/no record system with which I disagree in principle, since it places a high premium on the articulation and communication, in an extended series of statements, of sharable meanings, and since it makes perfectly clear that the final cause of the instructorial evaluation is to benefit the student. Evaluations derive primary justification from their importance to the intellectual growth and development of the student. As necessary aids to that intellectual development which has its proper culmination in competence, self-confidence, and independence, evaluations, it seems to me, beg for neither justification nor defense. Taking for granted, then, what is repeatedly emphasized in the arguments supporting credit/no record, I have concentrated, rather, on the extent to which various sign systems are intrinsically capable of conveying determinate meanings and on the value and reliability of the evidence upon which probability judgments, occasioned by the need to differentiate among students for largely extra-academic purposes, are based. Of necessity, such judgments will always be made, and I have maintained throughout (what I continue to believe) that the process of differentiation should reflect a painstaking examination and assessment of the achievement made manifest by students in a succession of endeavors over an extended period of time.

Now it is precisely on the matter of using evaluations as the primary resource of those responsible for differential judgments that I part company with the advocates of credit/no record, most of whom are willing to have those judgments based on letters of recommendation or general exams. Rather than reiterate the arguments of chapter 2 concerning the radical shortcomings of letters and general exams as sources of reliable

information and the fundamental weaknesses of the probability judgments based on them. I should now like to indicate briefly how credit/no record could be used in conjunction with a system of detailed evaluations.

I see no reason to believe that the integrity of the written evaluation proposal would be violated if at the end of the term the instructor sent to the registrar's office only a list of students who were to receive credit for the course and also submitted his evaluations on those students whom he could evaluate (not necessarily or usually every student in the class) to a kind of "placement" office, where they would be held until such time as the student decided to make use of them. The evaluations, in the aggregate, would be released only at the discretion of the students. What I have in mind is something analogous to the graduate student dossier (containing, normally, a record of courses taken–frequently with the grades–and letters of recommendation), which the student makes available to prospective employers. However, what distinguishes the folder of accumulated evaluations from the graduate student dossier is that it would contain no information of which the student was ignorant. (Graduate students rarely know how–i.e., in what terms–they are being commended and, consequently, are unable to defend themselves against either praise or blame.) When a student was notified that a decision was going to be made on his academic performance, or when he decided to apply for a position, a scholarship, or admission to a new program or school, he would, ideally, have at least the following three options: (1) he could request that the material in his folder be duplicated and then forwarded to the appropriate officer or committee; (2) he could send in place of the material in his folder a representative sampling of his work (as, for example, students today in the plastic and graphic arts submit photographs of their productions with their applications); or (3) he could submit for consideration both the folder and representative samples of his work. Of course, as long as the material is released only at the discretion of the students, there is no reason why the record of courses completed for credit and the folder of evaluations could not be stored in the registrar's office. Crucial to this proposal is the idea that the student record is personal, not public property. All sorts of agencies have a perfectly legitimate right to request information on students and to refuse to determine favorably in the absence of the requested information, but they have, it seems to me, no *right of access* to the personal records of students.

It would be possible to neglect the accumulated evidence relating to past achievement only when the job for which the student was applying required skills and talents totally unrelated to those necessary to academic work. In such cases, suitability for employment could be determined by a

personal interview and, perhaps, by performance on a special test designed (ideally, by the employer) to determine whether the student had any peculiar aptitude for the particular job. But such cases would, I think, be rare, primarily because the evaluations and samples of work reflect characteristics of mind and person that are relevant to successful performance on any job (whatever its nature and however remote its demands are from those involved in academic work). At any rate, it is important to remember that under this proposal the student would be responsible for deciding what kind of a case he would present to a prospective employer or to a selection board.

Such, then, in brief outline, is how I think credit/no record could be used in conjunction with the system of evaluative reports I have proposed. To my mind, the two proposals are mutually supportive, not naturally antagonistic.

To consider credit/no record in its full educational ramifications would take us well beyond the scope of this book, but before concluding this discussion, I should state that I look upon credit/no record as a tolerable, indeed desirable, interim arrangement, but one, however, which should ultimately be abandoned, since it continues to make accreditation, by implication at least, dependent upon a prescribed total of credit hours or units. Frankly, I see no reason why degree "requirements" could not be satisfied in a wide variety of ways. For example, a student—after taking a variety of courses, reading widely and assiduously on his own, recording detailed notes on his reading, and writing up the ideas and interpretations provoked by his study and inquiry—could present himself before a department or discipline review committee and ask for a decision on his competence in the area, broadly conceived, in which he had, in consultation with advisors, elected to work. The review committee would examine the material submitted (notes, interpretations, etc.), the available written evaluations of his work, and, perhaps, his performance on a general exam (oral, written, or both) and then decide whether in their judgment the student had demonstrated sufficient familiarity with an adequate range of primary and secondary material, with the basic scholarly or intellectual tools and resources of the discipline, and with the fundamental canons of demonstration and proof to justify the awarding of a degree or certificate to him. Under such a system the focus would be on what the student learned, not on the number of credits he accumulated, and such a system would be perfectly compatible with what is ultimately implicated in the demand for detailed evaluations—the replacement of quantitative with qualitative measures of competence and excellence.

Finally, the written evaluation system, as I have defined it, can easily

accommodate credit/no record; whether the accommodation is acceptable to the proponents of credit/no record, they, of course, must determine.

III

Most of the merits of the written report system have necessarily been delineated in the process of describing it or implied in the critique of grading procedures. However, a few benefits directly or indirectly following from the proposed reform should be spelled out in detail or reemphasized here. In the first place, instructors can transmit sharable meanings, not by means of grades or their various kinfolk, but by means of language. Among the immediate advantages accruing to a report system, then, is the sudden increase in the amount of useful information available to students and others. Students know where they are; they know, for example, what kinds of judgments are being made about their work and, consequently (if inferentially), what assumptions and standards are informing judgments.

Incidentally, when students have nothing but language to refer to, they must of necessity pay attention to what an instructor is writing in the margins and at the ends of their papers and exams. The university-wide adoption of the report system would undoubtedly halt the insane competition for grades and provide, I think, the necessary conditions for a healthy competition for excellence. Instead of competing with one another for grades, students would perhaps direct their energies toward meeting standards of excellence that they could respect. The enemies would no longer be fellow students, but unquestioned assumptions, inadequate evidence, and unfruitful hypotheses. Low grades, of course, should induce students to correct deficiencies, but very often students look only at their grades, not at accompanying comments, and determine only to do something different next time around. The intellectually disturbing fact is that grades are generally looked upon as the end of intellectual effort. When there are no grades, the end of effort will necessarily have to be differently construed. And clearly the student who writes a paper for a grade brings to his task a set of mind qualitatively different from that animating a student who knows that he can expect only comments on his essay. For the most part, grades only inspire in students a reluctance to work and a profound disdain for their skills, verbal and intellectual.

With or without grades many students, of course, will attempt to "psych-out" instructors and will continue to adjust their performances to conform with their notions of what they think instructors want, forsaking

their own ideas in an effort to placate a sometimes benevolent but potentially wrathful judge and succumbing to the silent injunction to obey the academic golden rule: "Do unto assignments what you suppose your instructors would reward you for doing unto them." However, students would perhaps be more willing than they currently are to trust their own powers and to substantiate original hypotheses if they were convinced that their arguments would be confronted head-on, with rejection of their views being contingent upon the emergence of unsuspected but damaging evidence or upon the patently superior explanatory power of counter-arguments. (When students are encouraged to look upon themselves as real participants in a genuine dialogue concerning the important and challenging problems of particular disciplines, they tend to behave like serious and responsible critics and scholars. The functional principle in all this is the "Doolittle" or "flower-girl" principle: "Treat a flower girl like a duchess, and she will behave like one." A student exercising intellectual muscles in the investigation of difficult problems does not tend to ask whether the activity can be justified; the activity is self-justifying, and satisfactory results provide the only reward necessary or meaningful.)

With the written report system, students would know, as I have said earlier, where they were, know, that is, how they were doing in relation to articulated or knowable standards and how well they were meeting specific intellectual challenges. If a student felt that his writings were regularly misunderstood or unfairly reviewed, he would, ideally, have an opportunity to submit to the registrar's office for placement in his record a written defense against certain charges or, perhaps, a sample of course work, which subsequent reviewers could use in determining, to their own satisfaction, the usefulness of the instructorial report or the degree of conformity existing between report and achievement. In all likelihood, students would seldom feel compelled to prepare materials for the purpose of self-defense, principally because it is reasonable to assume that instructors can justify their evaluations to students and that instructors and students can settle disagreements in conferences long before reports are filed with the registrar. Instructors who are convinced that their reports are both fair and justifiable cannot reasonably object to having their judgments subjected to particular scrutiny; and a student who refuses to be persuaded by either sound reasoning or irrefutable evidence will undoubtedly do his cause no good service by writing a defense in language that must inevitably expose the weaknesses of his case. However, when students are granted this recourse, they must inevitably feel less helpless than they presently do before the authority invested in the instructor to dispense or withhold favor. When the evaluation system allows students to

challenge or question the rulings of instructors, the threat implicit in absolute and final judgment is diminished, and students may no longer feel obliged to be wisely politic in how they handle their assignments. At any rate, with rebuttal built into the system, a mechanism is provided for the possible redress of real grievances and for limiting the influence of unfair or hostile judgments.

Even without the rebuttal provision, however, the written evaluation system by itself assures the student a fairer hearing than he is currently getting, for when an instructor writes out an evaluation, he inevitably exposes the bases of his judgments and the assumptions governing his remarks. Consequently, an evaluation having its origin in animosity or vindictiveness, in unfair or unsound principles, can possibly be seen for what it is, since reviewers are provided with an extended series of descriptive and normative terms. A grade, on the other hand, has an *ipse dixit* quality about it; it stands in silent and awful majesty on the page as an absolute, a timeless, unconditioned, irrefutable judgment. Of course, it is conditioned, but none of its determinants are known. A written evaluation, on the other hand, must necessarily reveal biases, prejudices and criteria of value, and when these are exposed, reviewers can determine—albeit according to their own standards—their worth. The written evaluation system does not guarantee the detection of judgments based on unfair, unsound, or inadequate principles, but it does provide the necessary conditions for detection, namely, an extended series of comments revealing directly or indirectly the priorities and emphases of particular instructors. Personal prejudice may still be disguised or go unnoticed, but the fact remains that, whereas grades can lie with impunity, written reports always run the risk of being charged with perjury or at least of being identified as hostile or unfair witnesses. Moreover, since reviewers see many reports on each student, they are in a position to recognize or to infer with some certainty which reports can and cannot be trusted and to know which they will or will not value highly. The system, in short, is a threat only to those instructors who suspect their motives or the value of their intellectual criteria and, of course, to those students who secure favor not so much by intellectual achievement as by chicanery or the more or less sophisticated ritual of academic courtship.

It is impossible here to discuss all the direct and adventitious benefits deriving from the implementation of a system of written evaluations, but a few additional ones can be noted, especially those that speak to problems raised earlier in this essay. In my opinion, the availability of written evaluations encourages reviewers, in and out of academia, to focus attention on what is most important, that is, academic performance in a

variety of courses over an extended period of time.

Additionally, the system has many of the virtues but none of the defects of the several selection criteria discussed earlier. For example, although written reports undoubtedly increase the bulk of student portfolios, they also do away, in effect, with the need for letters of recommendation, since what needs to be said about academic performance can be expressed in the evaluations by instructors, who are presumably qualified to discuss the work, at a time when they are most familiar with it. Under this system, then, instructors would avoid the painful and disagreeable task of trying to out-hyperbole one another and to recommend students whose work they may but dimly recall. And reviewers are thereby relieved of the burden of trying to discriminate among varying degrees of verbal excess.

Again, since reviewers most want to know how well a person meets the challenges of particular disciplines in his realized achievements (which expose the quality of his mind, the fruitfulness of his hypotheses, and perhaps the frequency with which he overcomes intellectual obstacles), they can neglect to consider scores achieved on standardized national exams, thereby saving students expense and a certain amount of needless anxiety. The student who by energy, perseverance, and assiduity manages to complete his assignments with distinction surely should not be eliminated from consideration because he scores poorly on a particular exam, however excellently that exam is constructed. When the nature and quality of accomplishment over an extended period of time can be known with some exactness, then, of course, it is pointless to make discriminations on the basis of a single test. In the so-called real world, advancement generally depends upon demonstrated competence in a succession of relevant and necessary tasks, not upon high scores on general exams. Written reports make selection on the basis of demonstrated on-the-job competence possible. Reviewers lose a convenient, if dubious, index to worth; however, inconvenience is more than compensated for by a net gain in reliable information that can be usefully employed in forecasting the nature of future performance.

One of the merits intrinsic to standardized national exams is anti-elitism (although many people would challenge these exams even under this heading). Ostensibly, at any rate, they discriminate according to achievement, not according to the prestige of the student's school or faculty sponsor. Written reports preserve this advantage, since they also are designed to differentiate among students on the basis of realized achievement. An individual report is valued in direct proportion to its capacity to account for phenomena; moreover, performance must be described and evaluated in terms that promote the transmission of sharable meanings,

and when an instructor, however great his reputation, can say very little about specific accomplishments, he must inevitably resort to indeterminate generalities or attribute to students qualities of which he has no direct knowledge, thereby running the risk of having his judgment falsified or devalued by virtue of its incongruity with judgments solidly based on detailed reference to specific achievements.

Under this system the settled inclination of reviewers may still be to emphasize the reports of nationally known professors at prestigious schools, but reviewers, like other men, focus on what is clearly and distinctly called to their attention and, consequently, on those reports which are most informative, which present the most convincing case, and which are at least partially confirmed by independent testimonies. As a result, every professor, regardless of rank and reputation, influences selection committees in proportion to the nature and quality of the information that he supplies. Presumably, those who pay closest attention to student work will be best able to describe and evaluate it in terms that reviewers can understand and according to criteria that are knowable, either directly or inferentially. Reviewers, of course, may or may not accept the legitimacy of the terms or accept the appropriateness of the values, but they must at least pay attention to the evidence submitted; only that information which has been casually prepared can they afford to dismiss casually. The injunction to instructors in all this is, "If you do not know the work, do not comment on it. If a student's work comes to you filtered through the consciousness of a graduate reader, either refuse to write an evaluation or submit—in his name—one that the graduate student has written." (An instructor giving a large lecture course and reading only the work of graduate students cannot, either physically or intellectually, evaluate the performances of all the undergraduates enrolled in the course.) But whoever submits the reports, the fact remains that the reviewers' attention will be controlled by the quality of the comments, not by the prestige of the writer or of the school.

By now, it should be clear that a change in evaluation procedures may have a profound impact on virtually all aspects of education, affecting matters ostensibly unrelated to evaluation. As suggested above, written reports would oblige professors teaching large lecture courses to omit reports on undergraduates altogether or to submit those of graduate readers. Of the two alternatives, I favor the former. Ideally, graduate students would still lead small discussion sections focusing on the content of the weekly lectures, but they would no longer be required to read exams or essays in order to determine how adept students are at recalling or approximating the ideas and language of the instructor. In my opinion,

lecture courses should be the burden of departments and should be offered regularly, with students attending as interest, need, or curiosity dictate. The lecture is undoubtedly a necessary, useful, and indispensable educational means for presenting information and suggesting interpretations and hypotheses, but students could demonstrate their understanding and control of the material delivered in lectures in their work for other courses. The point is that lecturers need not be required to take periodic test breaks. Students would attend lectures because they needed or wanted information, and acquisition of lecture material would be demonstrated by competent performance in those courses in which instructors could closely and personally examine their work. A student who, for one reason or another, could not enroll in all the lecture courses which interest him would be able to attend lectures in related or out of the way disciplines when the topic under discussion was particularly useful to his research, necessary to his understanding, or relevant to his long-range or broad (or, perhaps, whimsical) interests. Under the present credit system, an English student studying contemporary literature and its backgrounds might want to attend some of the lectures offered by the history department (or the physics department, for that matter) but might not be able or willing to sign up for the entire course, while taking at the same time several English courses. (In general, I think that we, as educators, should do as much as we can to make thinking of competence in terms of units and credits impossible.) Lectures could be variously organized (by historical sequence, by subjects, by periods, by problems, etc.), and since they would be the responsibility of departments—or perhaps ad hoc interdepartmental committees—individual lecturers would speak only on those matters in which they were particularly involved, retaining the role of lecturer only as long as they brought to it both interest and knowledge. Furthermore, the entire university would reap the benefits of this system if the lectures were publicly announced and open to the university community at large.

The universal adoption of the written report system of evaluation would also undoubtedly help to moderate the prevalence and occurrence of the first of the cardinal academic sins, printlust, to the commission of which every Ph.D. is prompted by universities, departments, colleagues, wives, husbands, pride, and so forth. The instructor who expends considerable time and energy evaluating student work will have little left over to devote to filling the, if not necessary, at least tolerable gaps in our knowledge or to throwing new and blinding light on subjects already illuminated to opacity. The emphasis here is on moderation, not on elimination; books and articles that a writer must eject from his system will find a way to bubble or roar their way into daylight, but surely there is something

pernicious about a system which requires every man, jack, and boy with a Ph.D., as a fundamental condition of his academic existence, to tear his guts apart, at who knows what ultimate expense, trying to write something for the sake of an associate professorship and deluding himself into thinking that once he reaches that academic Nirvana he can then write only what he wants to write, when and if he wants to write it. By the time he receives tenure, he has internalized the coercion, and he begins to look about him and wonder whether he is just a one-book man, who is not really pulling his share of the weight in the department.

Perhaps a revised evaluation system would prompt universities to recognize that publications are a desirable but unnecessary condition of employment and prompt instructors to insist that promotion and retention decisions focus on quality of mind and intellectual competence, however manifested in academic or professional work, and not primarily on publications or quality of mind as reflected in published material. (The guiding assumption here is not that those who publish teach poorly, whereas those who do not are excellent teachers; the fact is that many of the best teachers are also the most prolific writers. The essential point is that administrative focus should be on quality of mind, regardless of whether excellence is demonstrated in publications, teaching, or service to the academic or larger community. Of course, even universities willing to put more than token emphasis on teaching confront the problem of insufficient information and recognize that far too much weight must be given to hearsay evidence. Moreover, university instructors, for both sound and unsound reasons, have vigorously resisted any reforms that would permit colleagues to visit their classrooms regularly; consequently, much of what most needs to be known is inaccessible, except by means of the more or less reliable accounts of performance by colleagues and an unrepresentative portion of students.)

Obviously, a reform in the evaluation process will not overwhelm all the problems precluding a fair and rational assessment of teaching, but the written reports clearly make accessible the skills and powers that an instructor brings to one of his most important functions as a teacher. If the focus is on quality of mind, then it seems to me that the written reports would disclose information to hiring and retention committees fully as useful as that provided by publications. Whether the tenure system should be abandoned, along with other outdated and outmoded academic relics, is a question beyond the scope of this paper, but as long as tenure meetings continue to be held and as long as departments and schools are beginning to look seriously for reliable "teaching" data upon which to make frighteningly important decisions, one of the functions of the

written reports could be to furnish retention committees with important information about teaching competence—information revealing (sometimes directly and sometimes incidentally) an instructor's knowledge of subject matter and the particular habits and powers of his mind. (The process of assessing teaching without classroom visitations would be immensely refined, of course, if to the written reports were added the instructors' descriptions of the purposes and aims of their courses and evaluations of performance from all the students in their classes; it would take time to read through all of this material, but shortcuts to judgments are unacceptable when so much depends upon the decisions of the committee.)

Again, with written reports available to student advisers, advising might cease to be the rather perfunctory task it generally is today, involving little more than the signing of study lists prepared in advance by students. The reports would at least provide advisers with the sort of information that would enable them to do their job well, and, armed with this information, they would perhaps take an immediate and personal interest in the nature and direction of the students' academic programs, encouraging some to follow up their interests in graduate school, discouraging others from applying to schools to which they probably would not be admitted, and so on. At any rate, written reports would provide the necessary conditions for responsible advising.

Moreover, if the superintendents of scholarship and fellowship funds were supplied with written reports, they, too, would be forced to consider in detail the nature of a student's academic performance and to ask themselves whether they would be justified in removing support from a student whom instructors enjoin to take additional work. In all likelihood, these busy men would not want to devote long hours to the examination of these reports and might, consequently, reasonably elect to trust the validity of their original assessment and take the calculated risk of supporting students for extended periods of time without question, removing aid only when by choice or necessity the students withdrew from school. As a result, many students would be saved much useless anxiety and perhaps a good deal of sycophantic, hypocritical blubbering at the feet of instructors, upon whom so much of their future depends. The whole educational process would be humanized slightly if, collectively, we trusted our initial judgments more than we do and generally refused to ask those who bear our trust to prove repeatedly that we did not make foolish decisions or that they are not indolent wretches.

As stated earlier, it is impossible to predict all the consequences for education implicated in the proposed evaluation reform. Indeed, many

consequences cannot be known until the proposal is implemented. I have simply attempted to outline some of the possibilities that appear to me to be clearly and legitimately contained within the horizon constituted by the written evaluation system.

To those who might insist that the proposal, however attractive, cannot be implemented because its costs in time and energy are prohibitively high, the most direct response must be that I have written particular and summary reports on the work of my students for several years, without once asking them if they would object to my taking an incomplete in the courses on the grounds that I was unable to examine their work carefully in the time available to me. More important, slightly modified versions of the evaluation procedures that I have described in this paper have been adopted by several schools. The Santa Cruz campus of the University of California, for example, has been using a report system for several years now. The students at Santa Cruz know that their work is being closely examined, and the instructors know that they are involved in an evaluation process that is educationally sound and intellectually justifiable. Unquestionably, it takes more time to write reports than to assign grades, but when instructors are required to submit reports, they find the time to write them. Moreover, the written report system obliges an instructor to give no more time to evaluation than students have a right to ask of him. And to my knowledge, no one working within the system has complained about the extraordinary demands it makes upon his time. Clearly, what is being done can be done. (However, as long as the report system is adopted by only a relatively few schools of modest national reputation, the impact of the system on American education will be slight. If the system were implemented at five or six major universities, the reform would quickly become the uniform and standard policy of schools throughout the country, for the domino theory, however inadequate as an analysis of Asian political alignments, perfectly explains American higher education alignments. In higher education, as Harvard, Yale, etc., etc., go, so goes the nation. And if the system were adopted by a handful of "major" schools, administrators, selection committees, and personnel managers, would rapidly learn to cope with the problems that it incidentally created. Graduates from these schools would not cease to be admitted to law, medical, and graduate schools. What was being done at the so-called best schools would soon become the practice at most schools.)

To those who might object to written reports on the grounds that—in a period when increasing numbers of students are applying to graduate and undergraduate colleges—it would be impossible to give each student portfolio a careful examination, the immediate reply must be that only

those problems are solved which are confronted, and a system both intellectually and educationally sound should not be abandoned because it does not allow for the expeditious processing of data. The real burden is to solve the problems without undermining or violating the integrity of the system. And once administrators and reviewers accept the idea that the reports are a fact of educational life and are not subject to compromise, then ways of handling the reports are created and developed.

The problems themselves generate possible solutions. For example, on the graduate level admissions decisions might become the responsibility of departments. Instead of forming graduate admissions committees of four or five members, departments could ask every member to participate in the reviewing of credentials. Committees of three instructors of various ranks would review credentials and then make recommendations to the department. When all the committees of three had submitted recommendations, the entire department would meet as a committee to make final decisions. In general, too few faculty members today are responsible for decision making. If every department member were involved in making admissions decisions, no one would be overwhelmed with work, and every faculty member would have a hand in determining standards of admission and, to some extent, the nature of the graduate program. (Also, with widespread participation in decision making, faculty members would come to understand among themselves what kinds of students they wanted to attract and what they expected of them.)

Again, throughout this paper I have spoken of written reports, but surely the merits of the system would not be undermined if the reports were recorded not on paper but on tape. Department secretaries could either type up the reports submitted on tape or send the tapes directly to the registrar's office. The fact is that currently there are many technological devices that could be used to facilitate the transmission and handling of information, while preserving the content of the reports—tape recorders and duplicating machines of various kinds, for example. Whether my suggested solutions are feasible or not, however, is unimportant; it remains to be shown that the reports create insoluble problems. And the fact is that the problems have largely been solved by those schools which have adopted the system in one of its variations. (Moreover, for far too long educators have been willing to redefine or restructure their problems to fit the capabilities of technological devices or to conform to the ends generated by sound bookkeeping principles; ideally, the burden should be on technicians and economists to create structures and devices capable of serving the ends of education and of solving educational problems, without distorting or denying those ends and those problems.)

IV

The focus throughout this book has been on higher education, but what I have said clearly has implications for all levels of education; grades are as I have described them wherever they appear, and most of my arguments could certainly be directed to practices at the elementary and secondary school levels. Throughout, the silent reference has been to problems endemic to American education, and what is true of evaluation procedures at the university level is true, *mutatis mutandis*, of procedures at every level of American education.

The situation at San Francisco State calls attention to a general dilemma. All the efforts to explain away high grades cannot eradicate several inescapable facts: (1) grades are radically unsuited to the task of typifying academic behavior; (2) faculty members increasingly recognize the meaninglessness of grades and use them with less and less confidence; (3) students despise, but need grades, because they have social (and sometimes political), but not educational, value; and (4) reviewers regularly discount them and search about for alternative criteria of differentiation. The face-saving attempt to account for high grades provokes defensive teachers into extolling the superior intelligence of students and provokes college administrators into finding the locus of villainy in dissident teachers. Energy that could be employed in the development of new evaluation procedures is dissipated in the search for heroes and scapegoats; both parties indulge in their own peculiar forms of victimage. What this paper has tried to make clear is that the real villain is the grading system, and as long as it is in effect (in whatever variation), the only real victims will be those to whom grades are attached—students.

Bibliography

Aiken, Lewis R., Jr. "The Grading Behavior of a College Faculty." *Educational and Psychological Measurement* 23 (Summer 1963), pp. 319-322.

———. "Interdepartmental Variability and Student Expectations of College Grades." *Educational and Psychological Measurement* 24 (Winter 1964), pp. 823-829.

Anderhalter, O. F. "Developing Uniform Departmental Grading Standards in a University." *The Journal of Experimental Education* 31 (1962), pp. 210-211.

"Antioch Abolishes Grades." *School and Society* (November 9, 1968), p. 391.

Axelrod, Joseph. "The Creative Student and the Grading System." In *The Creative College Student: An Unmet Challenge*, edited by Paul Heist. San Francisco, California: Jossey-Bass, 1968.

———. "'What Do College Grades Mean?' A Survey of Practice at Four Institutions." In *College and University Teaching*, edited by H. A. Estrin and D. M. Goode, pp. 466-472. Dubuque, Iowa: Brown, 1964.

Bain, Philip T., et al. "Does Pass-Fail Encourage Exploration?" *College and University* 47 (February 1971), pp. 17-18.

Becker, Howard S.; Blanche Geer; and Everett C. Hughes. *Making the Grade: The Academic Side of College Life.* New York: Wiley, 1968.

Benson, Winston W. "Graduate Grading Systems." Washington, D.C.: Paper presented at the Council of Graduate Schools in the United States, 1969. ED 036 262, 7 pp.*

Birney, Robert C. "The Effects of Grades on Students." *The Journal of Higher Education* 35 (February 1964), pp. 96-98.

Bloom, Benjamin. "Learning for Mastery." *Evaluation Comment* 1 (May 1968).

Bramer, George R. "Grading and Academic Justice." *Improving College and University Teaching* 18 (Winter 1970), pp. 63-65.

Buchman, W. A. "Grading Systems across the Country . . ." Evanston, Illinois: Kendall College, Center for Curriculum Design, 1970.

Burke, Ronald J. "Student Reactions and Course Grades." *Journal of Experimental Education* 36 (Summer 1968), pp. 11-13.

Calhoon, R. P., and A. C. Reddy. "The Frantic Search for Predictors of Success." *Journal of College Placement* 28 (February-March 1968), pp. 54 ff.

Clark, Cecil D. "Competition for Grades and Graduate-Student Performance." *Journal of Educational Research* 62 (April 1969), pp. 351-354.

Clark, Edward L. "Reliability of Grade-Point Averages." *The Journal of Educational Research* 57 (April 1964), pp. 428-430.

Climo, James. "Pass/Fail at Longmeadow." *Clearing House* 43 (February 1969), pp. 341-343.

*Items in this bibliography which are available from ERIC (Educational Resources Information Center) have been assigned ED numbers. Reproductions of documents can be ordered in two forms: microfiche reproductions are available for 65 cents, regardless of the length of the text, and hard copy reproductions cost $3.29 for every hundred pages. Orders should specify the ED number, the type of reproduction desired, and the number of copies. They must be in writing and should be addressed to EDRS (ERIC Document Reproduction Service), P.O. Drawer 0, Bethesda, Maryland 20014. Payment must accompany orders under $10.00.

Cloger, Sheldon. "An Improved Grading System." *Education* 92 (September-October 1971), p. 95.

Cole, William G. "Breaking the Grade-and-Credit Mold." In *The Challenge of Curricular Change*, pp. 42-50. New York: College Entrance Examination Board, 1966.

Collins, Peter M. "Suggested Changes for Student Evaluation." *College Student Journal* 6 (February 1972), pp. 84-85.

Conway, J. A. "What Are We Rewarding?" *Phi Delta Kappan* 51 (October 1969), pp. 87-89.

Cromer, Ward. "An Empirical Investigation of Student Attitudes toward the Pass-Fail Grading System at Wellesley College." Philadelphia, Pennsylvania: Paper presented at Eastern Psychological Association, 1969.

Daiger, Richard. "Will Pass/Fail Pass the Test?" *Today's Education* 60 (November 1971), p. 24.

Dale, Wesley. "Concerning Grading and Other Forms of Student Evaluation." Washington, D.C.: Paper presented at the Council of Graduate Schools in the United States, 1969. ED 036 260, 6 pp.

Davis, Junius A. "What College Teachers Value in Students." *College Board Review* (Spring 1965), pp. 15-18.

Diederich, Paul B.; John W. French; and Sydell T. Carlton. *Factors in Judgments of Writing Ability*. Educational Testing Service Research Bulletin. Princeton, New Jersey: Educational Testing Service, 1961.

Disibio, Robert. "A Different Approach to Grading." *Education* 92 (November-December 1971), pp. 34-35.

Dressel, Paul L., and C. H. Nelson. "Testing and Grading Policies." In *Evaluation in Higher Education*, by Paul L. Dressel and Associates, pp. 227-252. Boston, Massachusetts: Houghton Mifflin, 1961.

Education at Berkeley: Report of the Select Committee on Education. Berkeley, California: University of California Press, 1968.

Elbow, Peter H. "More Accurate Evaluation of Student Performance." *The Journal of Higher Education* 40 (March 1969), pp. 219-230.

Elton, L. R. B. "The Assessment of Students—A New Approach." *Universities Quarterly* 22 (June 1968), pp. 291-301.

Ericksen, Stanford C. "Grades." *Memo to the Faculty*, No. 18. Ann Arbor, Michigan: University of Michigan, Center for Research on Learning and Teaching, 1966.

———. "Pass-Fail Grading." *Memo to the Faculty*, No. 22. Ann Arbor, Michigan: University of Michigan, Center for Research on Learning and Teaching, 1967.

Faust, Margaret S. *In Opposition to Conventional Grades.* ED 051 753, 8 pp.

Frechette, Ernest A. "Another Look at Marking." *Canadian Modern Language Review* 28 (January 1972), pp. 62-65.

Freeman, James T. "A Summary Progress Report on an Experimental Study of a Pass/No Report Grading System." San Bernardino, California: California State College at San Bernardino, 1969.

Friedenberg, Edgar Z. "The Real Functions of Educational Testing." *Change in Higher Education* 2 (1970), pp. 43-47.

Glazer, Nathan. "On Grading." In *Remembering the Answers: Essays on the American Student Revolt,* by Nathan Glazer, pp. 245-249. New York: Basic Books, 1970.

Gold, Richard M., et al. "Academic Achievement Declines under Pass-Fail Grading." *Journal of Experimental Education* 39 (Spring 1971), pp. 17-21.

Goldstein, Kenneth M., and Harvey A. Tilker. "Attitude: toward A-B-C-D-F and Honors-Pass-Fail Grading Systems." *Journal of Educational Research* 5 (November 1971), pp. 99-100.

———. "A Review of Grading System Practices in the Higher Education Institutions of New York State." Staten Island, New York: Richmond College, City University of New York, 1969.

Goodman, Paul. "In What Ways Does the Present Marking and Credit System Inhibit or Promote Learning?" In *Current Issues in Higher Education: Undergraduate Education,* edited by G. Kerry Smith, pp. 123-125. Washington, D.C.: Association for Higher Education, 1964.

Grades and Grading Systems. Higher Education Research Abstracts, No. 6. Hempstead, New York: Hofstra University, Center for the Study of Higher Education, March 1969.

"The Grading System at UCSC—A Critique." University of California, Santa Cruz, January 26, 1970. ED 037 174, 8 pp.

Haagen, C. Hess. "The Origins of a Grade." *The Journal of Higher Education* 35 (February 1964), pp. 89-91.

Hales, Loyde W., et al. "An Investigation of Some Aspects of the Pass-Fail Grading System." Paper presented at the meeting of the American Educational Research Association, 1971.

Hassler, William W. "Results of Pass-Fail Questionnaire Sent to Graduate School Deans." Indiana, Pennsylvania: Indiana University of Pennsylvania, 1969. ED 037 168, 4 pp.

Hepworth, Dean H. "The Predictive Validity of Admissions Criteria Used by the Graduate School of Social Work, University of Utah, to Select Students for the Fall Term, 1966." *Dissertation Abstracts* 29A (March 1969), pp. 3225-3226.

Hills, John R., and Marilyn B. Gladney. "Factors Influencing College Grading Standards." *Journal of Educational Measurement* 5 (Spring 1968), pp. 31-39.

Holt, John. "I Oppose Testing, Marking, and Grading." *Today's Education* 60 (March 1971), pp. 28-31.

Hoyt, Donald P. "College Grades and Adult Accomplishment." *Educational Record* 47 (Winter 1966), pp. 70-75.

———. "The Criterion Problem in Higher Education." In *Learning and the Professors,* edited by Ohmer Milton and Edward Joseph Shoben, Jr. Athens, Ohio: Ohio University Press, 1968.

———. "Rationality and the Grading Process." *Educational Record* 51 (Summer 1970).

———. "The Relationship between College Grades and Adult Achievement: A Review of the Literature." *ACT Research Report*, No. 7. Iowa City: American College Testing Program, 1965.

Humphreys, Lloyd G. "The Fleeting Nature of the Prediction of College Academic Success." *Journal of Educational Psychology* 59 (October 1968), pp. 375-380.

Hyman, Lawrence W. "Grades and Academic Standards." New York: Brooklyn College, City University of New York, 1969. ED 036 275, 6 pp.

Iadarola, Toni. "Grading the Graduate Student: A Student Perspective." Washington, D.C.: Paper read at the Council of Graduate Schools in the United States, 1969. ED 036 263, 7 pp.

Jencks, Christopher, and David Riesman. *The Academic Revolution.* New York: Doubleday, 1968.

Johansson, Charles B.; Jack E. Rossmann; and Sandra D. Sandell. "The S-U Grading Option." St. Paul, Minnesota: Macalester College, 1970.

Judy, Chester J. "A Proposed Technique for Improving Prediction of College Grades Using More Information from the High School Record." *Journal of Educational Measurement* 8 (February 1971), pp. 183-188.

Kelly, Samuel P., and Ralph Thompson. "Grading and the Nature of the Discipline." *The Journal of Higher Education* 39 (December 1968), pp. 517-518.

Kent, Leonard J. " 'Traditional' Graduate Grading and the 'Gold Star Syndrome.' " Council of Graduate Schools in the U.S., Washington, D.C., December 5, 1969. ED 037 152, 6 pp.

Kindsvatter, Richard. "Guidelines for Better Grading." *Clearing House* 43 (February 1969), pp. 331-334.

Kirby, B. C. "Three Error Sources in College Grading." *The Journal of Experimental Education* 31 (1962), pp. 211-218.

Kirschenbaum, Howard, et al. *Wad-Ja-Get? The Grading Game in American Education.* New York: Hart, 1971.

Klawitter, Robert. *Degrading Education. A Proposal for Abolishing the Grading System.* Washington, D.C.: Center for Educational Reform, National Student Association, 15 pp.

Kraft, Richard, and Gerald Lundquist. "Pass/Fail Grades Compared to Traditional Grading Procedures." *Journal of the Student Personnel Association for Teacher Education* 10 (September 1971), pp. 1-13.

Lacey, O. L. "How Fair Are Your Grades?" In *Higher Education in Engineering and Science,* edited by H. A. Estrin, pp. 348-352. New York: McGraw-Hill, 1963.

Lannholm, Gerald V. "Review of Studies Employing GRE Scores in Predicting Success in Graduate Study 1952-1967." Graduate Record

Examinations Special Report No. 68-1. Princeton, New Jersey: Educational Testing Service, 1968.

Lauter, Paul, and Florence Howe. "How the School System is Rigged for Failure." *The New York Review of Books* 14 (June 18, 1970), pp. 14-21.

Lavin, David E. *The Prediction of Academic Performance: A Theoretical Analysis and Review of Research.* New York: Russell Sage Foundation, 1965.

Law School Admission Test Council. "Statement on Pass-Fail Grading Systems." October 27, 1970. ED 045 043, 2 pp.

McGuire, Brian P. "Opinions Differ: The Grading Game." *Today's Education* 58 (March 1969), pp. 32-34.

Magoon, A. J., and R. Barker Bausell. "Comparing Pass/Fail Option against Traditional College Grading." *College Student Journal* 6 (February 1972), pp. 32-37.

Mannello, George. "Grades Must Go." *Educational Record* (Summer 1969), pp. 305-308.

Marshall, Max S. *Teaching without Grades.* Corvallis, Oregon: Oregon State University Press, 1968.

Merryman, Mary F., and Leo Kerwin. "A Three Semester Study of the Credit-No Credit Option in the College of Arts and Sciences, University of Missouri, Kansas City." University of Missouri, Kansas City, 1971. ED 051 750, 26 pp.

Pascal, Charles E. "Methods of Grading and Models of Teaching in Higher Education." *Educational Technology-Teacher and Technology Supplement* 11 (October 1971), pp. 59-60.

Midwest College Placement Association. "The Use of Personnel Tests and Biographical Data in the Selection of Recent College Graduates." West Lafayette, Indiana: The Association, 1964-65.

Miller, Richard H. "Students Show a Preparation Increase But No Increase in Grades Was Shown." *College and University* 45 (Fall 1969), pp. 28-30.

Miller, Stuart. *Measure, Number, and Weight: A Polemical Statement of the College Grading Problem.* Ann Arbor, Michigan: University of Michigan, Center for Research on Learning and Teaching, 1967. Also

Knoxville, Tennessee: University of Tennessee, Learning Research Center, 1967.

Milton, Ohmer. "Grades and Grading." *Teaching-Learning Issues*, No. 2. Knoxville, Tennessee: University of Tennessee, Learning Resources Center, 1966.

———. "Pass-Fail Grading." *Teaching-Learning Issues*, No. 5. Knoxville, Tennessee: University of Tennessee, Learning Resources Center, 1967.

———. " 'What It Is . . . I Measure I Do Not Know.' " *Educational Record* 49 (Spring 1968), pp. 160-165.

Paquette, J. W. "Student Credentials: What Elements Are Really Essential?" *Journal of College Placement* 26 (April-May 1966), pp. 34 ff.

"Pass-Fail Evaluation: Phase II: Questionnaire Analysis." Seattle, Washington: University of Washington, Office of Institutional Educational Research, April 27, 1970. ED 045 024, 30 pp.

Pass-Fail Study Committee. "Report to the Senate." Washington, D.C.: Phi Beta Kappa, 1969.

Pemberton, W. A. "The Grade Point Average: Snark or Boojum?" Newark, Del.: University of Delaware, September 1970. ED 047 009, 42 pp.

———. "A Recommendation For Revision of Student Evaluation Procedures." Newark, Delaware: University of Delaware, 1969.

Penner, Terry, and Tim Scanlon. *Comments and Counter Proposals to the Proposed Revision of the Grading System*. Princeton, New Jersey: Princeton University, March 1969.

Perry, Louis B. "College Grading: A Case Study and Its Aftermath." *Educational Record* 49 (Winter 1968), pp. 78-84.

Pitcher, Anita, and Robert Bosler. "Study of the 1969 UCSC Graduates." Santa Cruz, California: Office of the Chancellor, University of California, 1970.

Porter, Paul B. "Faculty Evaluation of Graduate Students." *Psychological Reports* 24 (1969), pp. 761-762.

Priest, R. F. "Attitudes to Competitive Grading in College as a Function of Aspiration, Anxiety, and Grades." Vancouver, British Columbia: Paper presented at the Western Psychological Association, 1969.

Quann, James C. "Pass/Fail Grading: Panacea or Placebo?" *Journal of the National Association of College Admissions Counselors* 16 (February 1972), pp. 11-13.

———. "Pass-Fail Grading: What Are the Trends?" Paper presented at the American Association of Collegiate Registrars and Admissions Officers, 1970.

———. "The Pass/Fail Option: Analysis of an Experiment in Grading." Athens, Ohio: American Association of Collegiate Registrars and Admissions Officers, April 27, 1971. ED 051 737, 18 pp.

Raimi, Ralph A. "Examinations and Grades in College." *AAUP Bulletin* 53 (September 1967), pp. 309-317.

Rasmussen, Glen R. "A Change in Grading Standards in Undergraduate Education Courses." *Journal of Higher Education* 40 (April 1969), pp. 311-317.

Rossmann, Jack E. "Graduate School Attitudes to S-U Grades." *Educational Record* 51 (Summer 1970), pp. 310-313.

Sgan, Mathew R. "The First Year of Pass-Fail at Brandeis University: A Report." *The Journal of Higher Education* 40 (February 1969), pp. 135-144.

Simon, Sidney B. "Down with Grades." *Today's Education* 58 (April 1969), p. 24.

———. "Grades Must Go." *School Review* 78 (May 1970), pp. 397-402.

Smallwood, Mary L. "An Historical Study of Examinations and Grading Systems in Early American Universities." *Harvard Studies in Education*, Vol. 24. Cambridge, Massachusetts: Harvard University Press, 1935.

Smith, Lawrence G. "Non-Punitive Grading in California Junior Colleges." El Cajon, California: Grossmont College, 1969. ED 035 409, 9 pp.

Sparks, David S. "Grading and Student Evaluation." Washington, D.C.: Paper presented at the Council of Graduate Schools in the United States, 1969. ED 036 261, 8 pp.

Spray, Cecil O. "Meaningful Grade Reporting." *Clearing House* 43 (February 1969), pp. 338-341.

Stallings, William M., and Elwood K. Leslie. "Student Attitudes toward

Grades and Grading." *Improving College and University Teaching* 18 (Winter 1970), pp. 66-68.

Stallings, William M.; H. Richard Smock; and Elwood K. Leslie. "The Pass-Fail Grading Option." *School and Society* 96 (March 16, 1968), pp. 179-180.

Teaf, H. M., Jr. "What Price Grades?" *The Journal of Higher Education* 35 (February 1964), pp. 100-103.

Thorndike, Robert L. "Marks and Marking Systems." In *Encyclopedia of Educational Research* (4th Edition), edited by R.L. Ebel, pp. 759-766. Toronto, Ontario: Macmillan, 1969.

Torbert, William R., and J. Richard Hackman. "Taking the Fun Out of Outfoxing the System: Yale University." In *The Changing College Classroom,* edited by Philip Runkel, Roger Harrison, and Margaret Runkel, pp. 156-181. San Francisco, California: Jossey-Bass, 1969.

University of California Committee on Educational Policy. "Report on Grading at the University of California, Santa Cruz," 1970.

University of California Committee on Grading. "Final Report and Recommendations." Berkeley, California, 1970.

Wardrop, James L. "Towards a Broader Concept of Educational Assessment." University of Illinois, Urbana: Center for Instructional Research and Curriculum Evaluation, February 14, 1972. ED 060 062, 28 pp.

Warren, Jonathan R. "College Grading Practices: An Overview." ERIC Clearinghouse on Higher Education, Washington, D.C., March 1971. ED 047 164, 29 pp.

Wilson, Kenneth M. "Increased Selectivity and Institutional Grading Standards." *College and University* 46 (Fall 1970), pp. 46-53.

Winkler, James E. "Grading." *Peabody Journal of Education* 46 (May 1969), pp. 350-351.

Wolfle, Dael. "Are Grades Necessary?" *Science* 161 (September 20, 1968), p. 1203.

Woodring, Paul. *"The Higher Learning in America: A Reassessment.* New York: McGraw-Hill, 1968.

Yelon, Stephen L. "An Alternative to Letter Grades." *Educational Forum* 35 (November 1970), pp. 65-70.

NATIONAL COUNCIL OF TEACHERS OF ENGLISH
1111 KENYON ROAD, URBANA, ILLINOIS 61801